ANALISIS SENTIMEN MASYARAKAT PENGGUNA MEDIA SOSIAL TERHADAP PROGRAM MAKAN BERGIZI GRATIS MENGGUNAKAN ALGORITMA *INDOBERT*

SKRIPSI

Oleh:
MUHAMMAD ALIF ATHALLAH
NIM. 220605110129

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026

ANALISIS SENTIMEN MASYARAKAT PENGGUNA MEDIA SOSIAL TERHADAP PROGRAM MAKAN BERGIZI GRATIS MENGGUNAKAN ALGORITMA *INDOBERT*

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
MUHAMMAD ALIF ATHALLAH
NIM. 220605110129

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2026

HALAMAN PERSETUJUAN

ANALISIS SENTIMEN MASYARAKAT PENGGUNA MEDIA SOSIAL TERHADAP PROGRAM MAKAN BERGIZI GRATIS MENGGUNAKAN ALGORITMA *INDOBERT*

SKRIPSI

Oleh:
MUHAMMAD ALIF ATHALLAH
220605110129

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 12 Juni 2026

Pembimbing I,	Pembimbing II,
Dr. M. Amin Hariyadi, M.T	Dr. Cahyo Crysdian, M.Cs
NIP. 19670118 200501 1 001	NIP. 19740424 200901 1 008

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M.Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

ANALISIS SENTIMEN MASYARAKAT PENGGUNA MEDIA SOSIAL TERHADAP PROGRAM MAKAN BERGIZI GRATIS MENGGUNAKAN ALGORITMA *INDOBERT*

SKRIPSI

Oleh:
MUHAMMAD ALIF ATHALLAH
NIM. 220605110129

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Tanggal: 24 Juni 2026

Susunan Dewan Penguji

Ketua Penguji	: Dr. Agung Teguh Wibowo Almais, M.T NIP. 19860301 202321 1 016	( )
Anggota Penguji I	: Roro Inda Melani, M.T, M.Sc NIP. 19780925 200501 2 008	( )
Anggota Penguji II	: Dr. M. Amin Hariyadi, M.T NIP. 19670118 200501 1 001	( )
Anggota Penguji III	: Dr. Cahyo Crysdian, M.Cs NIP. 19740424 200901 1 008	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: MUHAMMAD ALIF ATHALLAH
NIM	220605110129
Fakultas / Program Studi	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Analisis Sentimen Masyarakat Pengguna Media Sosial Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma *IndoBERT*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 26 Juni 2026
Yang membuat pernyataan,

MUHAMMAD ALIF ATHALLAH
NIM. 220605110129

MOTTO

"The Hardest Choices Requires The Strongest Wills"

HALAMAN PERSEMBAHAN

Alhamdulillahi rabbil 'alamin, segala puji bagi Allah Swt. atas segala rahmat, nikmat, dan karunia-Nya. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad saw., beserta keluarga, sahabat, dan seluruh umatnya. Atas izin dan pertolongan Allah Swt., penulis diberikan kemudahan dan kekuatan hingga dapat menyelesaikan skripsi ini.

Skripsi ini penulis persembahkan kepada kedua orang tua tercinta sebagai wujud bakti, cinta, kasih sayang, dan rasa terima kasih atas segala doa, pengorbanan, serta dukungan yang telah diberikan. Kehadiran dan doa mereka menjadi alasan utama yang menguatkan penulis untuk terus berjuang hingga mampu menyelesaikan skripsi ini.

Kepada saudara-saudara tercinta dan seluruh keluarga besar, penulis mempersembahkan karya sederhana ini sebagai bentuk rasa syukur dan terima kasih atas perhatian, motivasi, dukungan, serta doa yang selalu menyertai perjalanan penulis dalam menempuh pendidikan hingga terselesaikannya skripsi ini.

Kepada teman-teman dan seluruh pihak yang telah turut membantu penulis dalam menyelesaikan skripsi ini, terima kasih atas doa, dukungan, dan segala bentuk bantuan yang diberikan sehingga skripsi ini dapat terselesaikan dengan baik.

KATA PENGANTAR

Alhamdulillāhirabbil ‘ālamīn, segala puji bagi Allah *Subḥānahu wa Ta‘ālā*. yang telah melimpahkan rahmat, taufik, hidayah, serta karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul *“Analisis Sentimen Masyarakat Pengguna Media Sosial terhadap Program Makan Bergizi Gratis Menggunakan Algoritma IndoBERT”* dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada junjungan kita Nabi Muhammad *Shollallohu ‘Alaihi Wasallam*, beserta keluarga, sahabat, dan seluruh pengikutnya hingga akhir zaman. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nurdiana, M.Si., CAHRM., CRMP., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. M. Amin Hariyadi, M.T., selaku dosen pembimbing I dan Dr. Cahyo Crysdian, M.Cs, selaku dosen pembimbing II, yang telah memberikan bimbingan, arahan, masukan, serta motivasi kepada penulis selama proses penyusunan skripsi ini.

5. Dr. Agung Teguh Wibowo Almais, M.T., dan Roro Inda Melani, M.T,M.Sc selaku dosen penguji, yang telah memberikan evaluasi, saran, serta berbagai masukan yang berharga sehingga penelitian dan penulisan skripsi ini dapat tersusun dengan lebih baik. Terima kasih atas arahan yang diberikan dalam proses penyusunan skripsi.
6. Seluruh dosen Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang, penulis menyampaikan ucapan terima kasih atas ilmu yang telah di berikan selama masa perkuliahan.
7. Mama tercinta, Oty Rachmadhianty, yang selalu memberikan kasih sayang, doa, dukungan, dan semangat kepada penulis. Terima kasih karena selalu ada di setiap proses yang penulis lalui dan tidak pernah lelah memberikan motivasi hingga skripsi ini dapat diselesaikan. Terima kasih atas semua pengorbanan dan perhatian yang telah diberikan. Skripsi ini merupakan salah satu bentuk rasa syukur dan persembahan kecil dari penulis untuk Mama.
8. Papa tercinta, Iffan, yang telah memberikan doa, dukungan, nasihat, serta bekerja keras demi pendidikan dan masa depan penulis. Terima kasih atas segala pengorbanan dan keteladanan yang telah diberikan sehingga menjadi motivasi bagi penulis untuk terus berusaha dan menyelesaikan skripsi ini dengan sebaik-baiknya. Penulis berharap Papa selalu diberikan kesehatan, kekuatan, dan kebahagiaan.
9. Maheza Arkan Athallah, selaku adik penulis. Terima kasih karena telah menjadi salah satu motivasi bagi penulis untuk menyelesaikan pendidikan ini. Semoga suatu hari nanti kamu juga dapat mencapai semua impianmu dan meraih kesuksesan yang kamu cita-citakan.

10. Seluruh keluarga besar Suliah dan Achmad Eddy yang tidak dapat disebutkan satu per satu, terima kasih atas doa, dukungan, perhatian, dan semangat yang telah diberikan kepada penulis selama menempuh pendidikan.
11. Teman-teman terbaik penulis, VBT Boys, yaitu Muhammad Andrean Hidayatullah, Ahmad Fadhli Dzilikrom, Ki Ageng Nasrokh, Mohammad Zidan Alvaro, dan Muhammad Annas Darunnaja, yang telah menemani perjalanan penulis selama masa perkuliahan. Terima kasih atas kebersamaan, dukungan, bantuan, dan semangat yang selalu diberikan selama menjalani perkuliahan hingga proses penyusunan skripsi ini. Terima kasih untuk setiap cerita, diskusi, kerja sama, dan pengalaman yang telah dilalui bersama. Kehadiran kalian menjadi salah satu bagian yang membuat masa kuliah terasa lebih menyenangkan dan penuh kenangan.
12. Latif, Defa, dan Kevin yang telah menjadi teman berbagi cerita selama penulis berada di Bekasi. Terima kasih atas kebersamaan, obrolan, dan momen-momen yang telah dilalui bersama. Semoga kita semua diberikan kelancaran dalam meraih cita-cita dan tetap menjalin silaturahmi di mana pun berada.
13. Slebew Gaming dan The Weeknd, yang telah menjadi hiburan bagi penulis selama proses penyusunan skripsi melalui tayangan dan musik-musik nya.
14. Keluarga besar Teknik Informatika angkatan 2022, yang telah memberikan pengalaman berharga selama masa perkuliahan.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih terdapat berbagai keterbatasan yang memerlukan penyempurnaan di masa mendatang. Namun demikian, penulis meyakini bahwa proses belajar tidak memiliki batas akhir, sehingga setiap kritik dan saran yang membangun akan menjadi bekal berharga untuk terus berkembang. Bagi penulis, skripsi yang baik bukanlah skripsi yang sempurna, melainkan skripsi yang dapat diselesaikan dengan penuh tanggung jawab dan tepat pada waktunya. Oleh karena itu, penulis berharap skripsi ini dapat memberikan manfaat, baik bagi pengembangan ilmu pengetahuan maupun bagi pihak-pihak yang membutuhkan sebagai referensi untuk penelitian selanjutnya.

.

Malang, 26 Juni 2026

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Athallah, Muhammad Alif Athallah. 2026. **Analisis Sentimen Masyarakat Pengguna Media Sosial Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma *IndoBERT* .** Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. M. Amin Hariyadi, M.T. (II) Dr. Cahyo Crysdian, M.Cs.

Kata Kunci: Analisis Sentimen, *IndoBERT*, Klasifikasi Sentimen, Makan Bergizi Gratis

Program Makan Bergizi Gratis (MBG) merupakan salah satu program pemerintah yang menimbulkan berbagai tanggapan masyarakat di media sosial, baik berupa sentimen positif maupun negatif. Penelitian ini bertujuan untuk mengetahui kecenderungan sentimen masyarakat terhadap program Makan Bergizi Gratis (MBG) serta menganalisis kinerja algoritma IndoBERT dalam mengklasifikasikan sentimen teks berbahasa Indonesia. Penelitian dilakukan menggunakan data komentar media sosial dan diolah melalui tahapan preprocessing yang meliputi cleaning, normalisasi kata, tokenisasi, stopword removal, stemming, labelling, dan Random Oversampling. Hasil penelitian menunjukkan bahwa sentimen masyarakat terhadap program Makan Bergizi Gratis (MBG) cenderung didominasi oleh sentimen negatif dengan jumlah 2948 data, sedangkan sentimen positif berjumlah 1256 data. Selain itu, model IndoBERT mampu memberikan performa klasifikasi yang sangat baik dengan hasil terbaik diperoleh pada skenario 10 menggunakan rasio data 80:20, batch size 16, dan learning rate 0,00003 yang menghasilkan accuracy sebesar 95,85%, precision sebesar 94,42%, recall sebesar 97,46%, dan F1-score sebesar 95,91%.

ABSTRACT

Athallah, Muhammad Alif . 2026. **Sentiment Analysis of Social Media Users Toward the Free Nutritious Meal Program Using the IndoBERT Algorithm**. Undergraduate Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Supervisors: (I) Dr. M. Amin Hariyadi, M.T. (II) Dr. Cahyo Crysdian, M.Cs.

Key words: Sentiment Analysis, IndoBERT, Sentiment Classification, Free Nutritious Program

The Free Nutritious Meal Program (MBG) is one of the government initiatives that has generated various public responses on social media, both in the form of positive and negative sentiments. This study aims to determine the tendency of public sentiment toward the Free Nutritious Meal Program (MBG) and to analyze the performance of the IndoBERT algorithm in classifying Indonesian-language text sentiment. The research was conducted using social media comment data processed through several preprocessing stages, including cleaning, word normalization, tokenization, stopword removal, stemming, labelling, and Random Oversampling. The results indicate that public sentiment toward the Free Nutritious Meal Program (MBG) tends to be dominated by negative sentiment, with a total of 2,948 data entries, while positive sentiment accounts for 1,256 data entries. Furthermore, the IndoBERT model demonstrated excellent classification performance, with the best results achieved in Scenario 10 using an 80:20 data split ratio, a batch size of 16, and a learning rate of 0.00003, yielding an accuracy of 95.85%, precision of 94.42%, recall of 97.46%, and an F1-score of 95.91%.

البحث مستخلص

عطاالله، محمد أليف عطاالله. .2026 تحليل آراء مستخدمي وسائل التواصل الاجتماعي تجاه برنامج الوجبات الغذائية المجانية أطروحة تخرج .قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة IndoBERT. باستخدام خوارزمية د . (II) د .م .أمين هاريادي، ماجستير في الهندسة (I) :مولانا مالك إبراهيم الإسلامية الحكومية، مالانج .المشرفان كاهيو كريسديان، ماجستير في علوم الحاسوب.

الكلمات المفتاحية : تحليل المشاعر، IndoBERT، تصنيف المشاعر، الوجبات المغذية المجانية.

يعُد برنامج «الوجبات المغذية المجانية» (MBG) أحد البرامج الحكومية التي أثارت ردود فعل متنوعة من المجتمع على وسائل التواصل الاجتماعي، سواء كانت مشاعر إيجابية أو سلبية. تهدف هذه الدراسة إلى معرفة اتجاهات المشاعر العامة تجاه برنامج «الوجبات المغذية المجانية» (MBG) وتحليل أداء خوارزمية IndoBERT في تصنيف المشاعر في النصوص باللغة الإندونيسية. أجُريت الدراسة باستخدام بيانات التعليقات على وسائل التواصل الاجتماعي، وتم معالجتها من خلال مراحل ما قبل المعالجة التي شملت التنظيف، وتوحيد الكلمات، والتقطيع إلى رموز، وإزالة الكلمات غير المهمة، واستخلاص الجذور، والتصنيف، وأخذ العينات العشوائية الزائدة. أظهرت نتائج الدراسة أن المشاعر العامة تجاه برنامج الوجبات المغذية المجانية (MBG) تميل إلى أن تكون سلبية في الغالب، حيث بلغ عدد البيانات السلبية 2948، في حين بلغ عدد البيانات الإيجابية 1256. بالإضافة إلى ذلك، تمكن نموذج IndoBERT من تقديم أداء تصنيفي ممتاز، حيث تم الحصول على أفضل النتائج في السيناريو 10 باستخدام نسبة البيانات 80:20، وحجم الدفعة 16، ومعدل التعلم 0,00003، مما أدى إلى دقة بنسبة 95,85%، ودقة تحديد بنسبة 94,42%، واسترجاع بنسبة 97,46%، ودرجة F1 بنسبة 95,91%.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan media sosial telah mengubah pola interaksi dan cara masyarakat berbagi informasi. Saat ini, media sosial tidak hanya berfungsi sebagai alat komunikasi, tetapi juga telah menjadi ruang publik digital yang memfasilitasi penyampaian pendapat, kritik, serta dukungan terhadap isu-isu sosial maupun kebijakan publik.

Dalam perspektif Islam, penyampaian informasi dan pendapat perlu dilakukan secara bertanggung jawab dan didasarkan pada verifikasi kebenaran terhadap kebenaran infromasi tersebut, sebagaimana ditegaskan dalam Al-Qur'an Surah Al-Hujurat ayat 6 sebagai berikut:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَإٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْبِحُوا عَلَىٰ مَا فَعَلْتُمْ نَادِمِينَ

"Wahai orang orang yang beriman! Jika seorang yang fasik datang kepadamu membawa suatu berita, maka telitilah kebenarannya, agar kamu tidak mencelakakan suatu kaum karena kebodohan, yang akhirnya kamu menyesali perbuatanmu itu ." (QS. Al-Hujurat:6)

Berdasarkan tafsir ayat tersebut dijelaskan bahwa umat Islam diperintahkan untuk berhati-hati dalam menerima suatu informasi, terutama apabila informasi tersebut belum jelas kebenarannya. Apabila seseorang menerima berita tanpa melakukan verifikasi terlebih dahulu, maka hal tersebut dapat menimbulkan kesalahan dalam pengambilan sikap atau keputusan yang pada akhirnya dapat

merugikan pihak lain. Oleh karena itu, penerapan prinsip tabayyun atau verifikasi informasi menjadi sangat krusial agar masyarakat tidak mudah terpengaruh oleh informasi yang kebenarannya belum tervalidasi.

Nilai yang terkandung dalam ayat tersebut relevan dengan kondisi media sosial saat ini. Informasi yang beredar di media sosial sangat cepat dan sering kali disertai dengan berbagai opini yang beragam. Tanpa adanya upaya untuk memahami dan menganalisis informasi secara objektif, masyarakat berpotensi membentuk persepsi yang kurang tepat terhadap suatu isu atau kebijakan. Oleh karena itu, diperlukan suatu pendekatan yang mampu membantu memahami kecenderungan opini masyarakat secara lebih sistematis dan berbasis data.

Salah satu kebijakan yang memicu perdebatan opini di media sosial adalah program Makan Bergizi Gratis (MBG). Program bentukan pemerintah ini bertujuan untuk meningkatkan pemenuhan gizi masyarakat, dengan fokus utama pada kelompok siswa sebagai penerima manfaat.. Melalui program MBG, diharapkan kebutuhan gizi siswa dapat terpenuhi dengan lebih baik sehingga dapat mendukung kesehatan, konsentrasi belajar, serta kualitas sumber daya manusia di masa depan (Triningsih *et al*., 2025).

Implementasinya memunculkan berbagai respons dari masyarakat, terutama dari siswa dan konsumen yang merasakan langsung pelaksanaan program tersebut. Sebagian masyarakat memandang program ini sebagai langkah positif untuk meningkatkan kesejahteraan dan pemenuhan gizi para siswa.. Namun, sebagian lainnya menyampaikan kritik maupun kekhawatiran terkait kualitas makanan, distribusi program, serta pelaksanaan di lapangan. Perbedaan pandangan tersebut

memunculkan berbagai sentimen yang kemudian diekspresikan secara terbuka melalui media sosial.

Volume data opini masyarakat yang dihasilkan melalui media sosial, bersifat besar dan tidak terstruktur. Oleh karena itu, untuk memahami kecenderungan sentimen masyarakat secara objektif, diperlukan metode analisis yang mampu mengolah data teks dalam jumlah besar secara efektif. Analisis sentimen merupakan pendekatan dalam bidang *text mining* dan *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini berdasarkan kecenderungan sentimen. Pendekatan ini memungkinkan peneliti memperoleh gambaran umum persepsi publik tanpa harus melakukan analisis manual terhadap setiap teks (Waskito & Danang, 2019).

Namun, metode analisis sentimen konvensional yang mengandalkan representasi kata berbasis frekuensi memiliki keterbatasan dalam memahami konteks bahasa, terutama pada teks pendek seperti cuitan media sosial. Bahasa yang digunakan sering kali tidak baku dan sarat makna kontekstual, sehingga membutuhkan pendekatan yang mampu memahami hubungan antar kata dalam satu kalimat secara lebih mendalam.

Perkembangan teknologi deep learning, khususnya model berbasis transformer, telah memberikan kontribusi besar dalam peningkatan performa analisis teks. Salah satu model yang banyak digunakan adalah *Bidirectional Encoder Representations from Transformers* (BERT). Untuk konteks bahasa Indonesia, IndoBERT dikembangkan dan dilatih menggunakan korpus bahasa Indonesia sehingga mampu memahami konteks linguistik secara lebih mendalam

dibandingkan metode konvensional. Penggunaan IndoBERT dalam analisis sentimen diharapkan dapat menghasilkan klasifikasi sentimen yang lebih akurat dan relevan (Kusoema *et al*., 2025).

Berdasarkan uraian tersebut, penelitian ini mengajukan penggunaan IndoBERT sebagai metode utama dalam analisis sentimen. Pemilihan metode ini didasarkan pada kemampuannya dalam memahami konteks bahasa Indonesia secara lebih mendalam dibandingkan metode konvensional, sehingga diharapkan dapat menghasilkan klasifikasi sentimen yang lebih akurat. Melalui pendekatan ini, penelitian diharapkan dapat memberikan gambaran yang lebih objektif mengenai kecenderungan sentimen masyarakat terhadap kebijakan yang dibahas, sekaligus berkontribusi pada pengembangan analisis sentimen berbahasa Indonesia dengan pendekatan model transformer.

1.2 Pernyataan Masalah

Berdasarkan latar belakang yang telah diuraikan, pernyataan masalah dalam penelitian ini adalah bagaimana sentimen masyarakat pengguna media sosial program Makan Bergizi Gratis (MBG) serta bagaimana performa algoritma IndoBERT dalam mengklasifikasikan sentimen tersebut dan mengidentifikasi kecenderungan sentimen yang terhadap pelaksanaan program MBG.

1.3 Tujuan Penelitian

Mengetahui kecenderungan sentimen masyarakat dan menguji performa model IndoBERT dalam mengklasifikasikan teks berbahasa Indonesia.

1.4 Batasan Masalah

a. Data penelitian merupakan data sekunder yang diambil dari platform Kaggle
b. Data yang digunakan adalah komentar terhadap program Makan Bergizi Gratis
c. Sentimen dalam penelitian ini dibatasi pada dua kategori, yaitu sentimen positif dan negatif.
d. Data yang dianalisis hanya komentar yang ditulis dalam bahasa Indonesia dan tidak menggunakan konten non-teks

1.5 Manfaat Penelitian

Memberikan gambaran mengenai respons dan persepsi siswa atau konsumen terhadap program Makan Bergizi Gratis (MBG) yang tercermin dari opini publik di media sosial. Informasi mengenai kecenderungan sentimen tersebut dapat memberikan pemahaman mengenai bagaimana penerima manfaat menanggapi pelaksanaan program MBG di lapangan. Selain itu, hasil penelitian ini diharapkan dapat menjadi bahan evaluasi bagi pihak terkait, seperti pemerintah atau pemangku kebijakan, dalam mengevaluasi dan meningkatkan kualitas implementasi program MBG dengan mempertimbangkan respons yang berkembang di media sosial.

BAB II

STUDI PUSTAKA

Beberapa penelitian sebelumnya menunjukkan bahwa model BERT dan variannya, termasuk IndoBERT, memiliki performa yang baik dalam tugas analisis sentimen dibandingkan dengan metode machine learning konvensional.

Penelitian terdahulu oleh (Verena et al., 2021) menguji penggunaan metode *Naive Bayes* dan seleksi fitur RFFS untuk mengklasifikasikan pandangan masyarakat terkait kebijakan *New Normal*. Penelitian ini menggunakan 300 data tweet dan menerapkan skema k-fold cross validation. Hasil penelitian menunjukkan bahwa penerapan seleksi fitur mampu meningkatkan akurasi dari 62,6% menjadi 65,3%. Temuan ini menunjukkan bahwa metode konvensional masih memerlukan optimasi tambahan agar performanya meningkat, serta sensitif terhadap jumlah dan kualitas data yang digunakan.

Selain Naive Bayes, beberapa penelitian terdahulu juga menerapkan algoritma lain seperti *Support Vector Machine* (SVM). (Widowati et al., 2020) juga melakukan penelitian mengenai sentimen masyarakat terhadap Nadiem Makarim selaku Menteri Pendidikan dan Kebudayaan dengan memanfaatkan data dari Twitter. Penelitian tersebut membandingkan performa dua metode konvensional, yaitu *Naive Bayes* dan *Support Vector Machine* (SVM), dengan menerapkan pembobotan kata TF-IDF serta skema *k-fold cross validation*. Hasil analisis menunjukkan bahwa metode Naive Bayes memberikan performa yang lebih unggul dibandingkan dengan SVM.. Meskipun memberikan hasil yang cukup baik, penelitian ini memperlihatkan bahwa metode

konvensional masih bergantung pada representasi teks berbasis frekuensi kata dan belum mampu memahami konteks makna kalimat secara mendalam.

(Triningsih et al., 2025) melakukan penelitian yang menganalisis sentimen masyarakat terhadap program Makan Bergizi Gratis menggunakan basis data berupa 2.400 komentar dari media sosial X. Penelitian tersebut menggunakan algoritma Support Vector Machine (SVM) dan Random Forest untuk mengklasifikasikan sentimen masyarakat menjadi tiga kategori yaitu positif, negatif, dan netral. Algoritma SVM yang dikombinasikan dengan teknik *SMOTE* menghasilkan akurasi tertinggi mencapai 85,74%, lebih unggul daripada *Random Forest* yang hanya memperoleh akurasi sebesar 81,53%.

Penelitian (Pratama et al., 2025) membahas analisis sentimen publik terhadap Badan Pengelola Investasi (BPI) Danantara menggunakan algoritma IndoBERT pada platform media sosial X. Penelitian tersebut menggunakan 4.269 data tweet yang diperoleh melalui X API, yang kemudian dikelompokkan ke dalam tiga kategori sentimen positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa model IndoBERT mampu mencapai tingkat akurasi sebesar 97,71% serta efektif dalam mengolah data teks berbahasa Indonesia yang bersifat tidak terstruktur.

(Manoppo et al., 2025) melakukan analisis sentimen publik di media sosial mengenai kebijakan peningkatan Pajak Pertambahan Nilai (PPN) 12% di Indonesia dengan menerapkan model IndoBERT.. Data penelitian dikumpulkan dari beberapa platform media sosial, yaitu X, Instagram, dan TikTok, dengan total 2.581 data teks. Hasil analisis menunjukkan bahwa sentimen publik terhadap kebijakan

kenaikan PPN 12% didominasi oleh sentimen negatif. Model IndoBERT yang digunakan mampu mencapai akurasi sebesar 84,94% dan F1-score sebesar 84,37%, yang menunjukkan bahwa IndoBERT efektif digunakan untuk menganalisis respons masyarakat terhadap kebijakan publik berbasis opini di media sosial.

Selanjutnya, penelitian yang dilakukan (Al-kadzim, 2024) menganalisis perubahan sentimen publik di media sosial X terhadap isu konflik Palestina–Israel menggunakan model IndoBERT. Penelitian ini menunjukkan bahwa IndoBERT memiliki kemampuan yang baik dalam menangkap nuansa bahasa dan konteks opini publik yang kompleks, terutama pada isu-isu strategis dan sensitif. Hasil penelitian tersebut memperkuat temuan bahwa model berbasis transformer lebih unggul dibandingkan metode konvensional dalam analisis sentimen teks berbahasa Indonesia.

Selain itu, penelitian (Andhika et al., 2025) membahas analisis sentimen ulasan pengguna aplikasi Zoom menggunakan metode IndoBERT yang dikombinasikan dengan fitur ekstraksi GloVe. Penelitian ini menggunakan sebanyak 5.000 data ulasan yang diperoleh dari Google Play Store dan diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Untuk mengatasi permasalahan ketidakseimbangan kelas data, penelitian ini menerapkan teknik Random Oversampling. Hasil pengujian menunjukkan bahwa model gabungan IndoBERT dan GloVe mampu mencapai akurasi sebesar 91%, yang lebih tinggi dibandingkan penggunaan IndoBERT saja dengan akurasi 90%

Penelitian lain yang relevan dilakukan oleh (Jeevallucas et al., 2025) yang menganalisis sentimen pengguna terhadap akun resmi dompet digital DANA pada

platform X/Twitter menggunakan algoritma IndoBERT. Penelitian ini menggunakan 2.563 data tweet yang dikumpulkan selama periode Juli hingga November 2024 dan diklasifikasikan ke dalam kategori sentimen positif, netral, dan negatif. Proses penelitian meliputi tahap preprocessing, pembagian data secara stratifikasi, serta fine-tuning model IndoBERT. Hasil penelitian menunjukkan performa model yang sangat baik dengan precision sebesar 94%, recall 93%, F1-score 92%, dan akurasi 89%. Penelitian ini menegaskan keunggulan IndoBERT dalam menangkap konteks dan nuansa bahasa Indonesia pada data media sosial yang bersifat dinamis dan tidak terstruktur. Relevansi penelitian ini dengan skripsi yang dilakukan terletak pada kesamaan sumber data, yaitu platform media sosial serta penggunaan IndoBERT untuk menganalisis respons publik terhadap suatu layanan atau kebijakan, sehingga memperkuat dasar metodologis penelitian ini.

Tabel 2. 1 Penelitian Relevan

No	Nama Peneliti	Metode	Input	Output	Hasil
1	(Dalifah *et al*., 2024)	*IndoBERT*	Tweet pengguna platform X terkait BPI Danantara	Sentimen (positif, negatif, netral)	Penelitian ini memanfaatkan 23.623 data hasil pengembangan dari 4.269 tweet melalui proses augmentasi untuk melatih model IndoBERT. Hasilnya, model tersebut mampu mengidentifikasi sentimen publik terkait kebijakan BPI Danantara dengan akurasi

No	Nama Peneliti	Metode	Input	Output	Hasil
					mencapai 97,71%.
2	(Apryani *et al*., 2025)	SVM dan *Random Forest*	Tweet terkait program Makan Bergizi Gratis	Sentimen (positif, negatif, netral)	Hasil penelitian menunjukkan bahwa sentimen negatif mendominasi sebesar 46%. Model SVM dengan SMOTE memperoleh akurasi terbaik sebesar 85,74%, sedangkan Random Forest mencapai akurasi 81,53%.
3	(Farhan & Puspasari, 2025)	*IndoBERT*	Data media sosial terkait kebijakan PPN 12%	Sentimen (positif, negatif, netral)	Menunjukkan adanya dominasi sentimen negatif. Temuan ini didasarkan pada pengujian 2.581 data teks menggunakan model IndoBERT, yang berhasil mencatatkan tingkat akurasi sebesar 84,94% dan F1-score sebesar 84,37%.
4	(Ferdiyawan *et al*., 2025)	*IndoBERT*	Tweet pengguna terhadap akun resmi DANA di platform X	Sentimen (positif, negatif, netral)	Penelitian ini menggunakan 2.563 tweet dan menghasilkan performa model yang baik dengan precision 94%, recall 93%, F1-

No	Nama Peneliti	Metode	Input	Output	Hasil
					score 92%, dan akurasi 89%. IndoBERT terbukti mampu menangkap konteks bahasa Indonesia pada data media sosial
5	(Fauzan *et al*., 2025)	*Naive Bayes*	Tweet opini masyarakat mengenai kebijakan New Normal	Sentimen (positif, negatif, netral)	Menghasilkan akurasi rata-rata sebesar 62,6%, dan meningkat menjadi 65,3% setelah diterapkan seleksi fitur, yang menunjukkan adanya tantangan dalam memahami konteks bahasa pada metode konvensional.
6	(Imaddudin *et al*., 2025)	*Naive Bayes* dan *Support Vector Machine* (SVM)	Tweet pengguna Twitter dengan kata kunci "Nadiem Makariem", "Kemendikbud", dan "Pak Nadiem"	Sentimen (positif, negatif)	SVM memperoleh akurasi sebesar 85,47%. Hasil ini menunjukkan bahwa metode konvensional masih mampu memberikan performa yang baik, namun bergantung pada ekstraksi fitur dan pembobotan kata

No	Nama Peneliti	Metode	Input	Output	Hasil
7	(Pangestu *et al*., 2024)	*Naive Bayes* dan *Support Vector Machine* (SVM)	Tweet masyarakat mengenai program makan siang gratis	Sentimen (positif, negatif, netral)	Algoritma SVM menunjukkan performa dengan akurasi sekitar 75%.

2.1 Analisis Sentimen

Analisis sentimen merupakan salah satu metode dalam bidang *text mining* dan *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau pendapat seseorang terhadap suatu objek, peristiwa, atau isu tertentu. Analisis sentimen bertujuan untuk mengetahui kecenderungan sikap penulis teks, apakah bersifat positif, negatif, atau netral. Pendekatan ini banyak digunakan untuk mengolah data teks dalam jumlah besar yang bersumber dari berbagai media, seperti ulasan produk, berita daring, dan media sosial.

Dalam konteks media sosial, analisis sentimen menjadi penting karena media sosial merupakan ruang publik digital yang merepresentasikan opini masyarakat secara luas dan real-time. Setiap pengguna bebas mengekspresikan pandangan, kritik, maupun dukungan terhadap suatu isu, termasuk kebijakan publik. Oleh karena itu, analisis sentimen dapat digunakan sebagai alat untuk memahami persepsi dan respons masyarakat terhadap suatu kebijakan berdasarkan data opini yang disampaikan secara langsung oleh pengguna (R. A. Karim, 2024).

Analisis sentimen umumnya mengelompokkan opini ke dalam beberapa kategori sentimen. Klasifikasi yang paling umum digunakan adalah sentimen positif, negatif, dan netral. Sentimen positif menunjukkan adanya dukungan atau

penilaian baik terhadap suatu objek atau kebijakan, sentimen negatif mencerminkan ketidakpuasan atau penolakan, sedangkan sentimen netral menunjukkan opini yang bersifat informatif atau tidak mengandung kecenderungan emosional yang jelas. Pembagian kategori ini memudahkan peneliti dalam memahami pola dan kecenderungan opini masyarakat secara keseluruhan (Apriani & Gustian, 2019).

Pendekatan analisis sentimen dapat dilakukan dengan berbagai metode, mulai dari metode berbasis leksikon hingga metode berbasis pembelajaran mesin dan deep learning. Metode berbasis deep learning, khususnya model transformer seperti *IndoBERT*, memiliki keunggulan dalam memahami konteks bahasa secara lebih mendalam.

2.2 Text Mining

Text mining adalah proses pengolahan serta analisis data tekstual untuk mengekstrak informasi atau pola tersembunyi yang bermanfaat. Berbeda dari data numerik yang bersifat terstruktur, data teks pada umumnya tidak terstruktur sehingga membutuhkan tahapan pra-pemrosesan khusus sebelum dianalisis. Tujuan utama dari *text mining* adalah mentransformasikan teks mentah menjadi informasi bermakna yang dapat mendukung proses pengambilan keputusan.

Text mining merupakan proses pengolahan dan analisis data berupa teks untuk memperoleh informasi atau pola tertentu yang berguna. Berbeda dengan data numerik yang terstruktur, data teks umumnya bersifat tidak terstruktur sehingga memerlukan tahapan pengolahan khusus sebelum dapat dianalisis. Text mining bertujuan untuk mengubah data teks mentah menjadi informasi yang bermakna dan dapat digunakan sebagai dasar pengambilan keputusan (Sarimole, 2024).

Tahap ekstraksi fitur merupakan proses mengubah teks yang telah dibersihkan menjadi representasi numerik agar dapat diproses oleh algoritma komputasi. Pada tahap ini, setiap kata atau kalimat direpresentasikan dalam bentuk vektor dengan menggunakan metode tertentu. Representasi numerik ini sangat penting karena algoritma pembelajaran mesin dan deep learning hanya dapat bekerja dengan data berbentuk angka. Oleh karena itu, pemilihan metode ekstraksi fitur yang tepat sangat berpengaruh terhadap hasil analisis.

2.3 *Bidirectional Encoder Representations from Transformers (BERT)*

Bidirectional Encoder Representations from Transformers (BERT) merupakan model bahasa berbasis *deep learning* yang dirancang untuk memahami konteks tekstual secara mendalam. Dikembangkan oleh Google, model ini mampu mempelajari hubungan antar-kata dalam suatu kalimat dengan mempertimbangkan konteks dua arah baik dari kiri ke kanan maupun dari kanan ke kiri secara simultan.

Berbeda dengan model bahasa sebelumnya yang memproses teks secara searah, BERT menggunakan arsitektur transformer encoder yang memungkinkan model memahami makna suatu kata berdasarkan keseluruhan kalimat. Pendekatan dua arah ini menjadikan BERT lebih efektif dalam menangkap makna kontekstual, terutama pada kalimat yang memiliki struktur kompleks atau ambiguitas makna.

Arsitektur BERT dibangun berdasarkan mekanisme self-attention, yang memungkinkan model memberikan bobot perhatian pada setiap kata dalam kalimat. Dengan mekanisme ini, BERT dapat menentukan kata mana yang memiliki pengaruh lebih besar dalam memahami makna suatu kalimat. Proses ini membuat

BERT mampu mengenali hubungan semantik antar kata secara lebih akurat dibandingkan model berbasis urutan tradisional (Al-kadzim, 2024).

Proses pelatihan BERT bertumpu pada dua tugas utama, yakni *Masked Language Model* (MLM) dan *Next Sentence Prediction* (NSP). Melalui tugas MLM, sejumlah kata dalam kalimat sengaja disembunyikan agar model dapat memprediksi kata yang hilang tersebut berdasarkan konteks di sekitarnya. Sementara itu, NSP diterapkan untuk melatih kemampuan model dalam memahami keterkaitan antar-kalimat. Kombinasi kedua proses ini membekali BERT dengan pemahaman kebahasaan yang kokoh sebelum diimplementasikan pada tugas-tugas yang lebih spesifik (Fadilah et al., 2025).

Dalam penerapannya, BERT dapat digunakan kembali untuk berbagai tugas pemrosesan bahasa alami, seperti klasifikasi teks, analisis sentimen, dan pemahaman dokumen. Model ini tidak dilatih dari awal pada setiap penelitian, melainkan menggunakan model pra-latih (pre-trained model) yang kemudian disesuaikan dengan tugas tertentu. Pendekatan ini membuat BERT efisien dan efektif dalam berbagai aplikasi NLP.

2.4 *IndoBERT*

IndoBERT adalah salah satu model bahasa berbasis transformer yang dikembangkan secara khusus untuk menangani teks berbahasa Indonesia. Model ini merupakan adaptasi dari *Bidirectional Encoder Representations from Transformers* (BERT) yang dilatih menggunakan korpus bahasa Indonesia, seperti Wikipedia bahasa Indonesia dan berbagai sumber teks lokal lainnya. Dengan pelatihan pada data berbahasa Indonesia, IndoBERT mampu memahami karakteristik linguistik

bahasa Indonesia secara lebih baik dibandingkan model bahasa umum (Merdiansah & Ridha, 2024).

IndoBERT menggunakan pendekatan bidirectional, yaitu memproses konteks kata dari arah kiri ke kanan dan dari kanan ke kiri secara bersamaan. Pendekatan ini memungkinkan model memahami makna suatu kata berdasarkan keseluruhan konteks kalimat, bukan hanya berdasarkan kata sebelumnya atau sesudahnya saja. Kemampuan tersebut sangat penting dalam analisis sentimen, terutama pada data media sosial yang sering mengandung bahasa tidak baku, singkatan, dan konteks makna yang beragam.

Secara arsitektur, IndoBERT menggunakan encoder transformer yang terdiri dari beberapa lapisan *self-attention* dan *feed-forward neural network*. Setiap teks masukan terlebih dahulu diubah menjadi representasi embedding sebelum diproses oleh model. Embedding pada IndoBERT merupakan gabungan dari tiga komponen utama, yaitu token embedding, segment embedding, dan positional embedding, yang dirumuskan sebagai berikut:

$$E_{input} = E_{token} + E_{segment} + E_{position} \tag{2.1}$$

Keterangan:

E_{token} : representasi token atau sub-kata hasil tokenisasi

$E_{segment}$: digunakan untuk membedakan segmen kalimat

$E_{position}$: posisi token dalam urutan kalimat

Dalam tugas klasifikasi teks, IndoBERT menggunakan token khusus [CLS] sebagai representasi keseluruhan kalimat. Representasi token [CLS] ini digunakan

sebagai masukan utama untuk menentukan kelas sentimen dari suatu teks. Proses klasifikasi dapat dirumuskan sebagai berikut:

$$\hat{y} = \text{softmax}(W \cdot h_{CLS} + b) \tag{2.2}$$

Keterangan:

h_{CLS} : vektor representasi token [CLS]

W : matriks bobot klasifikasi

b : bias

$\hat{y}$: probabilitas prediksi kelas sentiment

Fungsi softmax digunakan untuk menghasilkan probabilitas setiap kelas sentimen, yaitu positif, negatif, dan netral. Untuk mengukur kesalahan prediksi selama proses pelatihan, digunakan fungsi categorical cross-entropy loss. Fungsi loss ini berperan dalam mengoptimalkan model agar prediksi sentimen semakin mendekati label yang sebenarnya. Secara matematis, fungsi loss dirumuskan sebagai berikut:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{2.3}$$

Keterangan:

C : jumlah kelas sentimen

y_i : label sentimen sebenarnya

$\hat{y}_i$: probabilitas prediksi kelas ke – i

Dalam penelitian ini, IndoBERT digunakan sebagai model utama untuk melakukan analisis sentimen terhadap data teks yang bersumber dari platform X.

Pemilihan IndoBERT didasarkan pada kemampuannya dalam menghasilkan representasi bahasa Indonesia yang kontekstual dan akurat, sehingga diharapkan mampu memberikan hasil klasifikasi sentimen yang representatif.

2.5 *Early Stopping*

Early stopping merupakan salah satu teknik yang digunakan dalam proses pelatihan model machine learning untuk mencegah terjadinya overfitting. Overfitting terjadi ketika model terlalu baik dalam mempelajari data latih, namun kurang mampu melakukan generalisasi terhadap data baru atau data uji.

Selama proses training, model akan terus mengalami penurunan nilai training loss. Namun, tidak selalu diikuti oleh penurunan validation loss. Dalam beberapa kondisi, validation loss justru mulai meningkat setelah beberapa epoch. Hal ini menandakan bahwa model mulai menghafal data latih dan kehilangan kemampuan untuk mengenali pola secara umum (H. F. Karim & Wibowo, 2025).

Early stopping bekerja dengan cara menghentikan proses pelatihan model secara otomatis ketika performa pada data validasi tidak lagi mengalami peningkatan. Biasanya, kriteria yang digunakan adalah ketika validation loss tidak membaik dalam beberapa epoch berturut-turut. Dengan demikian, model akan berhenti pada titik terbaik sebelum mengalami overfitting.

BAB III

DESAIN DAN IMPLEMENTASI SISTEM

3.1 Akuisisi Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari platform Kaggle. Kaggle dipilih sebagai sumber data karena menyediakan berbagai dataset yang dapat dimanfaatkan untuk keperluan penelitian di bidang analisis data dan pembelajaran mesin. Dataset yang digunakan dalam penelitian ini berisi kumpulan komentar masyarakat yang berkaitan dengan program Makan Bergizi Gratis (MBG).

Data yang tersedia dalam dataset tersebut berupa teks komentar yang memuat opini, tanggapan, maupun pandangan masyarakat terkait program Makan Bergizi Gratis (MBG). Dataset yang digunakan memiliki total 4.899 data komentar yang kemudian digunakan sebagai sumber data dalam penelitian ini.

Meskipun dataset telah tersedia dalam bentuk kumpulan komentar, data yang diperoleh masih berupa data mentah yang belum siap untuk langsung dianalisis. Data tersebut masih mengandung berbagai elemen yang tidak relevan untuk proses analisis teks, seperti tanda baca, simbol, URL, serta penggunaan kata tidak baku yang umum ditemukan dalam percakapan di media sosial. Oleh karena itu, sebelum digunakan dalam proses analisis sentimen, data terlebih dahulu melalui tahap pembersihan dan pengolahan data (preprocessing) agar teks menjadi lebih terstruktur dan sesuai untuk diproses oleh model. Jumlah data yang digunakan dalam penelitian ini dapat dilihat pada Tabel 3.1.

Tabel 3. 1 Sampel Data

No	Sumber Data	Jenis Data	Jumlah Data
1	Kaggle	Komentar terkait program MBG	4899
Total			**4899**

Data juga diperiksa dari sisi kelengkapan teks. Cuitan yang kosong, terpotong, atau tidak dapat diproses lebih lanjut dikeluarkan dari data penelitian. Setelah melalui proses tersebut, diperoleh kumpulan data teks yang lebih terfokus dan siap untuk diproses lebih lanjut pada tahap pengolahan data. cuitan direpresentasikan dalam bentuk teks pada kolom cuitan, kemudian diberi label sentimen yang terdiri dari dua kategori, yaitu positif dan negatif. Berdasarkan dataset yang ditampilkan pada tabel 3.1, data berupa kumpulan komentar dari media sosial.

Tabel 3. 2 Contoh Komentar

No	Komentar	Label
1	Daerah pedesaan anak sangat antusias kalo sekolah elit jelas makanannya turun kasta	Positif
2	MBG untuk kesehatan kenapa ambil banyak porsi dana pendidikan untuk MBG ini kalo MBG untuk kesehatan tapi banyak yang keracunan malah makin aneh dong	Negatif
3	Setuju buat apa makanan gratis tapi toh tidak sehat juga malah keracunan buang duit triliunan yang bener tidak bikin maju	Negatif
4	Makanan beracun gratis dialihkan jadi sekolah gratis sarjana pendidikan benar gratis dan fasilitas lebih penting dari makan mulu silahkan anak makan sendiri dengan memberikan pekerjaan untuk orang tua nya masing	Positif
5	Makanan berbakteri itu tidak selalu basi saya pernah kena muntaber sebelumnya saya tidak makan makanan basi pun	Positif
6	Programnya tidak perlu kata gratis walaupun gratis karena nanti tertanam dibenak anak tentang gratisan	Positif

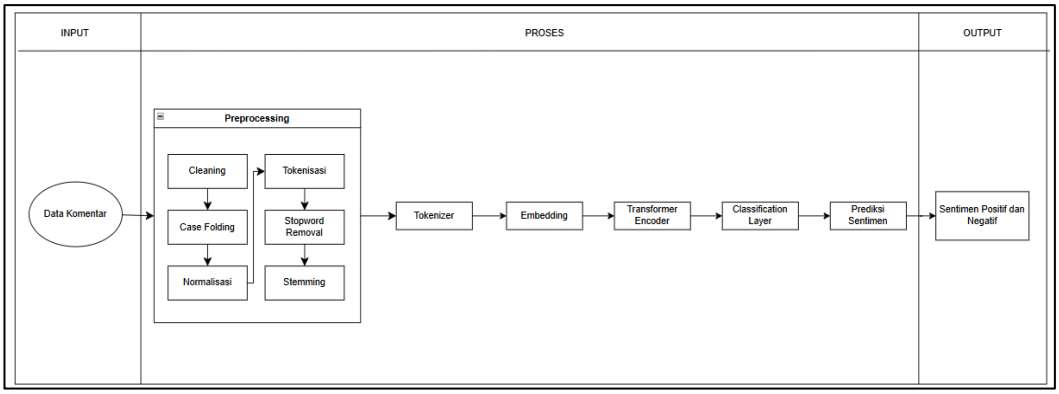
No	Komentar	Label
7	Maaf Pak bnyak menu MBG nggak sesuai dengan uang yang dtetapkan jadi menu seperti dpangkas jadi rbu Pak	Negatif
8	MBG itu cuma jadi lahan basah korupsi kalo tidak di kawal dengan benar ibaratnya dari atas daging sampe bawah jadi ikan teri	Negatif
9	Makanan beracun gratis, dan ade gua salah satu korbannya	Negatif
10	Karena kebutuhan gizi setiap siswa berbeda jadi harus dihitung kalorinya dan siswa yang alergi beberapa makanan harus diperhatikan meskipun jadi ribet tapi agar tepat sasaran	Positif
11	Program makan gratis bagus oke tapi pelaksanaan nya bagaimana? Setahu saya banyak program bagus tapi sering kali pelaksanaan kacau, di Konoha seperti itu biasanya...	Negatif
12	Program makan gratis bagus oke tapi pelaksanaan nya bagaimana? Setahu saya banyak program bagus tapi sering kali pelaksanaan kacau, di Konoha seperti itu biasanya...	Positif
13	Mau sebagus apapun program nya, kalau eksekusi nya jelek ya jadi jelek, tinggal bagaimana cara kepala mengatur semua nya apakah tutup mata atau mau memperbaiki.	Negatif

3.2 Desain Sistem

Desain sistem pada penelitian ini menggambarkan alur pengolahan data teks hingga proses klasifikasi sentimen. Sistem dirancang untuk mengolah cuitan mentah dari media sosial menjadi data yang siap digunakan dalam pelatihan dan pengujian model analisis sentimen. Secara umum, sistem terdiri dari beberapa proses utama, yaitu pembersihan data, preprocessing teks, pelatihan model, serta evaluasi hasil klasifikasi.

Hasil klasifikasi dari masing-masing model kemudian dievaluasi pada tahap evaluasi model. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall,

dan f1-score, serta confusion matrix untuk melihat kinerja model dalam mengklasifikasikan sentimen positif dan negatif secara lebih rinci.



Gambar 3. 1 Desain Sistem

3.2.1 *Data Cleaning*

Tahap cleaning data bertujuan untuk membersihkan cuitan dari elemen-elemen yang tidak relevan dan dapat mengganggu proses analisis sentimen. Data yang diperoleh dari media sosial masih bersifat mentah dan mengandung berbagai komponen tambahan seperti username, tautan, tanda baca, simbol, serta karakter khusus yang tidak memiliki kontribusi terhadap makna sentimen. Pada penelitian ini, proses cleaning data difokuskan pada penghilangan bagian-bagian teks yang tidak diperlukan. Username pengguna dihapus untuk menghindari bias dan menjaga fokus analisis pada isi cuitan. Selain itu, tautan seperti URL dan alamat situs web juga dihilangkan karena tidak berpengaruh terhadap sentimen yang disampaikan dalam teks. Tanda baca, karakter khusus, serta elemen non-alfabet turut dibersihkan agar teks menjadi lebih sederhana dan mudah diproses. Proses ini juga mencakup penghapusan karakter HTML dan spasi berlebih yang sering muncul pada data hasil crawling. Dengan demikian, setiap cuitan diubah menjadi teks yang lebih bersih dan konsisten.

3.2.2 *Preprocessing*

Tahap *preprocessing* dilakukan untuk mempersiapkan data teks agar lebih terstruktur dan mudah dipahami oleh model analisis sentimen. Setelah melalui proses cleaning data, teks cuitan masih memiliki variasi penulisan dan bentuk kata yang dapat memengaruhi hasil klasifikasi apabila tidak ditangani dengan baik. Oleh karena itu, *preprocessing* diperlukan untuk menyederhanakan teks tanpa menghilangkan makna utama yang terkandung di dalamnya. Pada tahap ini, setiap cuitan diproses melalui serangkaian teknik pengolahan teks yang umum digunakan dalam analisis sentimen.

Preprocessing bertujuan untuk menyeragamkan bentuk teks, mengurangi noise, serta menyesuaikan data dengan kebutuhan model pembelajaran mesin dan model bahasa yang digunakan dalam penelitian. Proses *preprocessing* yang diterapkan mencakup beberapa langkah, antara lain penyeragaman huruf (*case folding*), pembersihan lanjutan terhadap karakter yang tidak diperlukan, normalisasi kata tidak baku, pemisahan teks menjadi token, penghapusan kata-kata umum yang tidak memiliki nilai sentimen, serta proses stemming untuk mengubah kata ke bentuk dasarnya.

3.2.3 *Case Folding* dan *Text Cleaning*

Tahap ini dilakukan untuk menyeragamkan bentuk teks serta menghilangkan elemen-elemen yang tidak diperlukan sebelum data dianalisis lebih lanjut. Pada tahap ini, seluruh teks cuitan diubah menjadi huruf kecil untuk menghindari perbedaan makna akibat variasi penggunaan huruf besar dan kecil. Dengan penyeragaman ini, kata yang memiliki bentuk sama namun ditulis dengan

huruf berbeda akan dianggap sebagai satu kata yang sama. Selain itu, dilakukan pembersihan teks terhadap berbagai komponen yang tidak berkontribusi pada analisis sentimen. Username pengguna, tautan URL, dan alamat situs web dihapus karena tidak mengandung informasi sentimen. Tanda baca, karakter khusus, serta elemen HTML juga dibersihkan agar tidak mengganggu proses pengolahan data.

```
import re

def clean_text(text):

text = re.sub(r'rt\s+', '', text)

text = re.sub(r'http\S+', '', text)

text = re.sub(r'@\w+', '', text)

text = re.sub(r'#\w+', '', text)

text = re.sub(r'[^a-zA-Z\s]', '', text)

text = re.sub(r'\s+', ' ', text).strip()

return text

df['text'] = df['text'].apply(clean_text)
```

Pada proses ini, teks difilter sehingga hanya menyisakan huruf alfabet, sedangkan angka dan karakter non-alfabet dihilangkan. Spasi berlebih serta karakter tunggal yang tidak memiliki makna juga dihapus untuk menghasilkan teks yang lebih ringkas dan konsisten. Dengan demikian, setiap cuitan diubah menjadi bentuk teks yang lebih bersih dan siap untuk diproses pada tahap selanjutnya.

Tabel 3.3 Proses Case Folding

Sebelum	**Setelah**
Mending kasih Susu saja dan Roti sudah Cukup bang dari pada di kasih Makanan yg harus di Olah dulu ðŸ‘‰	mending kasih susu saja dan roti sudah cukup bang dari pada di kasih makanan yang harus di olah dulu
Makanan beracun gratis, dan ade gua salah satu korbannya	makanan beracun gratis dan ade gua salah satu korbannya

Sebelum	Sesudah
Di kota tempat saya tinggal, tampak masih belum siap program ini. Satu kota hanya ada dua dapur, jadi masih belum merata.	di kota tempat saya tinggal tampak masih belum siap program ini satu kota hanya ada dua dapur jadi masih belum merata
Program makan gratis bagus oke tapi pelaksanaan nya bagaimana? Setahu saya banyak program bagus tapi sering kali pelaksanaan kacau, di Konoha seperti itu biasanya...	program makan gratis bagus oke tapi pelaksanaan nya bagaimana setahu saya banyak program bagus tapi sering kali pelaksanaan kacau di konoha seperti itu biasanya

Berdasarkan hasil yang ditunjukkan pada Tabel 3.3, proses cleaning dan case folding dilakukan untuk membersihkan teks dari berbagai elemen yang tidak diperlukan serta menyeragamkan bentuk huruf pada setiap komentar.

3.2.4 Normalisasi

Normalisasi dilakukan untuk menyamakan bentuk kata yang tidak baku atau bersifat slang menjadi kata baku yang memiliki makna yang sama. Pada data cuitan media sosial, sering ditemukan penggunaan singkatan, bahasa tidak formal, atau variasi penulisan kata yang berbeda-beda. Jika tidak dinormalisasi, perbedaan tersebut dapat menyebabkan model menganggap kata yang sebenarnya sama sebagai kata yang berbeda.

```
def normalize_text(text):
    if isinstance(text, str):
        words = text.split()
        normalized = [normalisasi_dict[word] if word in normalisasi_dict else word
for word in words]
        return " ".join(normalized)
    else:
        return text
df['text'] = df['text'].apply(normalize_text)
```

Pada penelitian ini, proses normalisasi dilakukan dengan memanfaatkan kamus kata tidak baku yang disimpan dalam bentuk file CSV. Kamus tersebut berisi pasangan kata tidak baku dan padanannya dalam bentuk kata baku. Setiap token pada cuitan akan dicocokkan dengan kamus normalisasi, kemudian kata yang ditemukan akan diganti dengan bentuk bakunya, sedangkan kata yang tidak terdapat dalam kamus akan tetap digunakan sebagaimana adanya

Tabel 3.4 Proses Normalisasi

Sebelum	Setelah
anakku dah kls sd tapi belum dpt mbgpdhl d jatim	anakku sudah kelas sd tapi belum dapat mbg padahal di jatim
aku gak setuju bukan karena program nya gagal, tapi gak tepat aja waktunya	aku tidak setuju bukan karena program nya gagal tapi tidak tepat saja waktunya
Maaf Pak bnyak menu MBG nggak sesuai dengan uang yang dtetapkan jadi menu seperti dpangkas jadi rbu Pak	maaf pak banyak menu mbg tidak sesuai dengan uang yang ditetapkan jadi menu seperti dipangkas jadi ribu pak

Berdasarkan hasil pada Tabel 3.4, tahap normalisasi dilakukan untuk memperbaiki kata-kata yang tidak baku atau menggunakan singkatan yang sering muncul dalam komentar pengguna. Pada teks media sosial, pengguna sering menuliskan kata dengan bentuk yang disingkat atau tidak sesuai dengan kaidah bahasa Indonesia, seperti penggunaan kata “udh”, “dpt”, atau “gk”. Bentuk penulisan seperti ini dapat menyulitkan sistem dalam mengenali makna kata secara tepat.

3.2.5 Tokenisasi

Teks komentar yang sebelumnya masih berupa kalimat utuh kemudian diuraikan menjadi bagian-bagian yang lebih sederhana berupa kata. Pemecahan ini

membantu sistem memahami isi teks secara lebih terstruktur, karena setiap kata dapat dianalisis secara terpisah tanpa bergantung pada bentuk kalimat aslinya. Dalam penelitian ini, pemisahan kata dilakukan dengan pendekatan tokenisasi berbasis pola. Pendekatan ini memungkinkan teks dipisahkan secara fleksibel berdasarkan karakter pembatas seperti spasi, sehingga kata-kata yang muncul dalam cuitan dapat dikenali dengan lebih baik, termasuk kata yang berasal dari bahasa tidak formal atau hasil normalisasi sebelumnya.

Melalui proses ini, satu cuitan tidak lagi diperlakukan sebagai satu kesatuan teks, melainkan sebagai sekumpulan token yang merepresentasikan makna dari cuitan tersebut. Token-token inilah yang selanjutnya digunakan dalam proses penghapusan stopwords dan stemming. Dengan demikian, tokenisasi berperan penting dalam membentuk struktur data teks yang siap digunakan dalam analisis sentimen dan pelatihan model klasifikasi.

Tabel 3.5 Proses Tokenisasi

Sebelum	Setelah
tapi kan program ini karna tingginya angka stunting di indo susu dan roti hanya akan membuat obesitas	['tapi', 'kan', 'program', 'ini', 'karna', 'tingginya', 'angka', 'stunting', 'di', 'indo', 'susu', 'dan', 'roti', 'hanya', 'akan', 'membuat', 'obesitas']
kalau soal makanan gratis bagi saya sih bagus untuk murid yang memang uangnya tidak ada sama yang buat nabung	['kalau', 'soal', 'makanan', 'gratis', 'bagi', 'saya', 'sih', 'bagus', 'untuk', 'murid', 'yang', 'memang', 'uangnya', 'tidak', 'ada', 'sama', 'yang', 'buat', 'nabung']
program sudah lama berjalanmasa baru sekarang ada kejadian keracunan pemerintah juga tidak mungkin diam	['program', 'sudah', 'lama', 'berjalanmasa', 'baru', 'sekarang', 'ada', 'kejadian', 'keracunan', 'pemerintah', 'juga', 'tidak', 'mungkin', 'diam']
mbg sudah tepat pak saya sebagai ibu yang punya anak sekolah sangat terbantu	['mbg', 'sudah', 'tepat', 'pak', 'saya', 'sebagai', 'ibu', 'yang', 'punya', 'anak', 'sekolah', 'sangat',

terimakasih pak prabowo sehat selalu bapak presiden	'terbantu', 'terimakasih', 'pak', 'prabowo', 'sehat', 'selalu', 'bapak', 'presiden']
mbg sangat bermanfaat sukses selalu pak prabowo subianto	['mbg', 'sangat', 'bermanfaat', 'sukses', 'selalu', 'pak', 'prabowo', 'subianto']

3.2.6 *Stopwords Removal*

Setelah teks dipecah menjadi kumpulan token, langkah berikutnya berfokus pada penyaringan kata-kata yang terlalu umum dan tidak memiliki pengaruh langsung terhadap penentuan sentimen. Kata-kata seperti kata sambung, kata ganti, atau istilah umum sering muncul dalam cuitan, namun keberadaannya tidak memberikan informasi penting terkait sentimen positif atau negatif.

```
stop_words = stopwords_df.iloc[:, 0].tolist()

stop_words = set(stop_words)


print(list(stop_words)[:20])

df['tokens'] = df['tokens'].apply(
    lambda words: [word for word in words if word not in stop_words]
)
```

Dalam penelitian ini, daftar stopwords disusun dengan menggabungkan stopwords bawaan bahasa Indonesia serta stopwords tambahan yang sering muncul pada percakapan di media sosial. Stopwords tambahan tersebut mencakup kata slang, kata serapan, serta istilah informal yang sering digunakan dalam cuitan namun tidak memiliki makna sentimen yang kuat.

Token yang termasuk dalam daftar stopwords kemudian dihilangkan dari data. Proses ini bertujuan untuk menyederhanakan teks dengan mempertahankan kata-kata yang lebih bermakna dan relevan terhadap analisis sentimen. Dengan

berkurangnya kata-kata umum, fokus analisis dapat diarahkan pada kata-kata yang benar-benar mencerminkan opini atau emosi pengguna.

3.2.7 *Stemming*

Kata-kata yang tersisa masih dapat memiliki berbagai bentuk turunan, seperti awalan, akhiran, maupun imbuhan lainnya. Perbedaan bentuk kata ini dapat menyebabkan sistem menganggap kata yang sebenarnya memiliki makna sama sebagai kata yang berbeda. Oleh karena itu, dilakukan proses stemming untuk menyederhanakan kata ke bentuk dasarnya.

Stemming dilakukan dengan menggunakan algoritma stemming bahasa Indonesia, sehingga proses penghilangan imbuhan disesuaikan dengan kaidah morfologi bahasa Indonesia. Setiap token hasil penghapusan stopwords diproses untuk menghilangkan prefiks, sufiks, maupun konfiks yang melekat pada kata, sehingga kata-kata dengan makna serupa dapat direpresentasikan dalam satu bentuk yang sama.

Penerapan stemming membantu mengurangi variasi kosakata dalam data teks tanpa menghilangkan konteks utama dari cuitan. Dengan berkurangnya variasi bentuk kata, model dapat lebih fokus pada makna inti yang terkandung dalam teks, bukan pada perbedaan struktur kata. Hasil dari proses stemming berupa kumpulan kata dasar yang lebih konsisten dan terstandarisasi. Data inilah yang kemudian digunakan sebagai representasi akhir teks sebelum masuk ke tahap pelatihan model.

3.3 *IndoBERT*

Data yang digunakan pada tahap ini merupakan data teks yang telah melalui seluruh rangkaian preprocessing, mulai dari cleaning data hingga stemming,

sehingga teks berada dalam bentuk yang lebih bersih, konsisten, dan representatif. Data hasil preprocessing ini kemudian digunakan sebagai masukan dalam proses pelatihan model analisis sentimen.

Proses pelatihan model memanfaatkan beberapa library utama, yaitu Pandas untuk pengelolaan dan manipulasi data, Scikit-learn untuk pembagian data latih dan data uji, serta Transformers dari Hugging Face untuk penggunaan model bahasa pra-latih IndoBERT. Selain itu, PyTorch digunakan sebagai backend utama dalam proses pelatihan model berbasis deep learning.

Sebelum pelatihan dilakukan, data terlebih dahulu diberi label numerik dengan memetakan label sentimen NEGATIF menjadi nilai 0 dan POSITIF menjadi nilai 1. Pemetaan ini diperlukan agar label dapat diproses secara komputasional oleh model klasifikasi. Selanjutnya, data dibagi menjadi data latih dan data uji dengan skema pembagian 80% data latih dan 20% data uji menggunakan fungsi train_test_split. Pembagian ini bertujuan untuk menguji kemampuan model dalam mengklasifikasikan data yang belum pernah digunakan selama proses pelatihan.

```
from transformers import BertTokenizer


model_name = "indobenchmark/indobert-base-p1"
tokenizer = BertTokenizer.from_pretrained(model_name)
```

Model yang digunakan dalam penelitian ini adalah IndoBERT dengan arsitektur indobenchmark/indobert-base-p1. Model ini dipilih karena telah dilatih sebelumnya menggunakan korpus bahasa Indonesia dalam jumlah besar, sehingga memiliki kemampuan yang baik dalam memahami konteks dan struktur bahasa

Indonesia, khususnya pada teks pendek seperti cuitan media sosial. Proses tokenisasi teks dilakukan menggunakan AutoTokenizer dari library Transformers dengan pengaturan panjang maksimum teks agar sesuai dengan kebutuhan model.

Data teks yang telah ditokenisasi kemudian dikonversi ke dalam bentuk dataset menggunakan kelas dataset berbasis PyTorch. Dataset ini berfungsi sebagai penghubung antara data hasil preprocessing dan proses pelatihan model, sehingga data dapat diproses secara batch dan efisien selama proses training. Melalui tahapan ini, model dilatih untuk mengenali pola-pola sentimen pada data teks dan mempelajari perbedaan karakteristik antara sentimen positif dan negatif.

3.4 *IndoBERT*

Uji coba awal dilakukan untuk memastikan bahwa model IndoBERT dapat berfungsi dengan baik sebelum diterapkan pada keseluruhan dataset penelitian. Pada tahap ini, digunakan sebagian data, sebagai langkah awal untuk mengevaluasi kesiapan model, alur preprocessing, serta konfigurasi pelatihan yang digunakan.

Data yang digunakan pada uji coba awal merupakan data teks yang telah melalui seluruh tahapan preprocessing, sehingga berada dalam bentuk yang bersih dan siap diproses oleh model. Setiap data cuitan telah diberi label sentimen positif atau negatif sesuai dengan hasil pelabelan yang dilakukan sebelumnya.

Pada tahap ini, digunakan sebagian data, sebagai langkah awal untuk mengevaluasi kesiapan model, alur preprocessing, serta konfigurasi pelatihan yang digunakan.

Data yang digunakan pada uji coba awal merupakan data teks yang telah melalui seluruh tahapan preprocessing, sehingga berada dalam bentuk yang bersih dan siap diproses oleh model. Setiap data cuitan telah diberi label sentimen positif atau negatif sesuai dengan hasil pelabelan yang dilakukan sebelumnya.

Teks cuitan kemudian diproses menggunakan tokenizer IndoBERT. Tokenisasi bertujuan untuk mengubah teks menjadi token-token yang dikenali oleh model. Setiap token kemudian diubah menjadi ID token dan dilengkapi dengan attention mask agar model mengetahui token mana yang perlu diproses.

Token yang telah diubah menjadi ID numerik kemudian diproses oleh lapisan encoder IndoBERT. Pada tahap ini, IndoBERT mempelajari hubungan antar kata dalam kalimat menggunakan mekanisme self-attention. Secara konseptual, hubungan antar kata dihitung menggunakan persamaan:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \qquad (3.5)$$

Melalui proses ini, kata seperti “cepat panas” dan “ragu” memperoleh bobot perhatian yang lebih besar karena memiliki keterkaitan kuat dengan ekspresi sentimen negatif.

Output dari token khusus [CLS] digunakan sebagai representasi keseluruhan kalimat. Representasi ini kemudian diteruskan ke lapisan klasifikasi untuk menentukan probabilitas sentimen. Misalnya, model menghasilkan nilai logit sebagai berikut:

- Logit negatif = 2.15
- Logit positif = 0.85

Nilai tersebut kemudian dihitung menggunakan fungsi softmax:

$$P(negatif) = \frac{e^{2.15}}{e^{2.15} + e^{0.85}} = 0.79$$

$$P(positif) = \frac{e^{0.85}}{e^{2.15} + e^{0.85}} = 0.21$$

Karena probabilitas sentimen negatif lebih besar, maka hasil prediksi model untuk cuitan tersebut adalah sentimen negatif (0). Misalkan label sebenarnya dari cuitan tersebut adalah negatif (0), maka nilai cross-entropy loss dihitung sebagai berikut:

$$\mathcal{L} = -\log(P(\text{negatif})) = -\log(0.79) = 0.235$$

Setelah proses pelatihan awal selesai, model diuji menggunakan data uji pada tahap uji coba awal. Hasil prediksi dibandingkan dengan label sebenarnya untuk menghitung nilai akurasi, presisi, *recall*, dan *F1-score*. Evaluasi ini bertujuan untuk memastikan bahwa model mampu mengenali pola sentimen dengan benar dan tidak mengalami kesalahan mendasar sebelum diterapkan pada pengujian utama. Misalkan dari 300 data uji coba awal, diperoleh hasil sebagai berikut:

- *True Positive* (TP) = 70
- *True Negative* (TN) = 180
- *False Positive* (FP) = 20
- *False Negative* (FN) = 30

Maka perhitungan metrik evaluasi adalah:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{70 + 180}{300} = 0.83$$

$$\text{Presisi} = \frac{TP}{TP + FP} = \frac{70}{70 + 20} = 0.78$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{70}{70 + 30} = 0.70$$

$$\text{F1-score} = \frac{2 \times (0.78 \times 0.70)}{0.78 + 0.70} = 0.74$$

BAB IV
UJI COBA DAN PEMBAHASAN

4.1 Skenario Uji Coba

Tahap ini dilakukan untuk menilai kemampuan model dalam mengklasifikasikan sentimen secara tepat dan konsisten. Pada penelitian ini, evaluasi dilakukan dengan membandingkan hasil prediksi model terhadap label sentimen sebenarnya pada data uji. Proses evaluasi ini penting untuk memastikan bahwa model tidak hanya mampu menghasilkan prediksi, tetapi juga memiliki kinerja yang dapat dipertanggungjawabkan secara ilmiah.

Untuk memperoleh gambaran performa model secara menyeluruh, digunakan beberapa metrik evaluasi, yaitu akurasi, presisi, recall, dan F1-score. Setiap metrik memiliki fungsi yang berbeda dan saling melengkapi dalam menilai kinerja model klasifikasi.

Akurasi digunakan untuk melihat seberapa besar proporsi data yang dapat diklasifikasikan dengan benar oleh model secara keseluruhan. Metrik ini memberikan gambaran umum tentang tingkat keberhasilan model, namun tidak selalu mencerminkan performa yang sebenarnya apabila distribusi data antar kelas tidak seimbang. Akurasi dihitung dengan membandingkan Jumlah prediksi yang benar terhadap seluruh data yang diuji, dengan rumus sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \qquad (3.1)$$

Presisi berfungsi untuk mengukur tingkat ketepatan model dalam memprediksi suatu kelas sentimen. Presisi menunjukkan seberapa besar prediksi yang dihasilkan oleh model untuk suatu kelas benar-benar sesuai dengan label sebenarnya. Metrik ini penting untuk mengetahui apakah model cenderung menghasilkan prediksi yang keliru atau berlebihan terhadap suatu kelas sentimen. Presisi dapat dihitung dengan rumus:

$$Presisi = \frac{TP}{TP+FP} \tag{3.2}$$

Recall digunakan untuk menilai kemampuan model dalam mengenali seluruh data yang sebenarnya termasuk ke dalam suatu kelas sentimen. Metrik ini menunjukkan seberapa baik model dalam menangkap data yang relevan tanpa melewatkannya. Recall menjadi penting ketika tujuan penelitian adalah meminimalkan kesalahan akibat data yang tidak terdeteksi oleh model. Rumus recall adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \tag{3.3}$$

F1-score digunakan sebagai metrik penyeimbang antara presisi dan recall. Metrik ini memberikan gambaran performa model yang lebih stabil, terutama ketika terdapat perbedaan distribusi jumlah data pada masing-masing kelas sentimen. F1-score menggabungkan presisi dan recall ke dalam satu nilai sehingga dapat mencerminkan kinerja model secara lebih adil. F1-score dihitung menggunakan rumus:

$$F1\text{-}score = \frac{2\times(Presisi\times Recall)}{Presisi+Recall} \quad (3.4)$$

Skenario pengujian disusun untuk mengetahui kemampuan model IndoBERT dalam mengklasifikasikan sentimen cuitan ke dalam kategori positif dan negatif. Pengujian dilakukan setelah proses pelatihan model selesai, dengan menggunakan data uji yang telah dipisahkan sebelumnya agar hasil pengujian dapat menggambarkan performa model secara objektif.. Berikut rincian skenario pengujian tertera pada Tabel 4.1

Tabel 4.1 Skenario Uji Coba

Skenario	Data Latih	Data Uji	Batch Size	Learning Rate
Skenario 1	60%	40%	16	0.00002
Skenario 2	60%	40%	16	0.00003
Skenario 3	60%	40%	32	0.00002
Skenario 4	60%	40%	32	0.00003
Skenario 5	70%	30%	16	0.00002
Skenario 6	70%	30%	16	0.00003
Skenario 7	70%	30%	32	0.00002
Skenario 8	70%	30%	32	0.00003
Skenario 9	80%	20%	16	0.00002
Skenario 10	80%	20%	16	0.00003
Skenario 11	80%	20%	32	0.00002
Skenario 12	80%	20%	32	0.00003

Pada skenario uji coba, setiap komentar pada data uji diproses menggunakan tahapan yang sama seperti data latih, mulai dari preprocessing hingga tokenisasi

sesuai dengan format input IndoBERT. Model kemudian menghasilkan prediksi sentimen untuk setiap cuitan berdasarkan pola yang telah dipelajari selama proses pelatihan.

Hasil prediksi model dibandingkan dengan label sentimen sebenarnya untuk menghitung nilai performa. Evaluasi dilakukan menggunakan beberapa metrik, yaitu accuracy, precision, recall, dan f1-score, guna melihat ketepatan dan konsistensi model dalam mengklasifikasikan sentimen. Selain itu, confusion matrix digunakan untuk menunjukkan distribusi prediksi benar dan salah pada masing-masing kelas sentimen.

4.2 Hasil Uji Coba

Pengujian dilakukan menggunakan tiga perbandingan data, yaitu 80:20, 70:30, dan 60:40. Pada setiap skenario, model dilatih menggunakan data latih yang telah ditentukan dengan batch size 16, 32 dan learning rate 0.00002, 0.00003 kemudian dievaluasi menggunakan data uji. Evaluasi performa model dilakukan dengan menggunakan confusion matrix serta metrik evaluasi seperti accuracy, precision, recall, dan F1-score.

4.2.1 Skenario Model 1

Pada skenario ini, model diuji menggunakan pembagian data sebesar 60:40, yaitu 60% data sebagai data latih dan 40% sebagai data uji. Model telah melewati proses *fine-tuning* selama 2 epoch dengan parameter batch size sebesar 16 dan learning rate sebesar 0.00002.

Epoch	Training Loss	Validation Loss
1	0.272081	0.310097
2	0.141402	0.253342

Gambar 4.1 Hasil Training Skenario 1

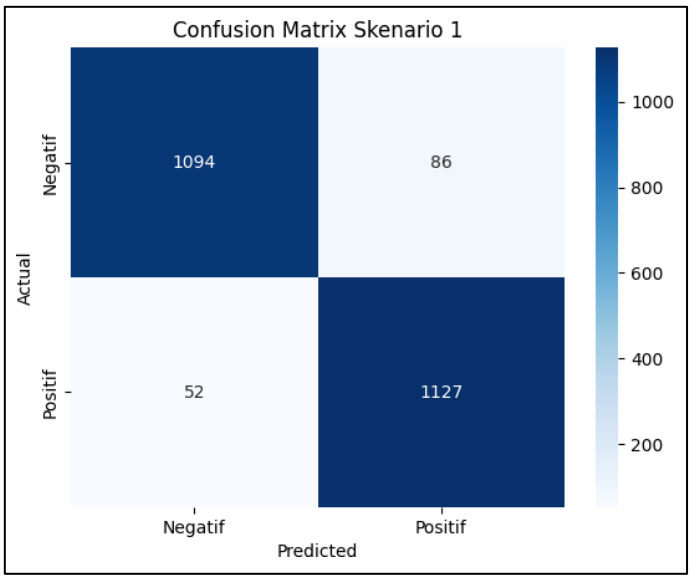
Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* dan *validation loss* yang cukup baik. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.272081 dan *validation loss* sebesar 0.310097. Kemudian pada *epoch* kedua nilai *training loss* menurun menjadi 0.141402 dan *validation loss* menurun menjadi 0.253342.

Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.2.

Tabel 4.2 Hasil Evaluasi Skenario 1

Akurasi (%)	**Presisi (%)**	**Recall (%)**	**F1-Score (%)**
94.15	92.91	95.59	94.23

Berdasarkan hasil pengujian pada data uji, model memperoleh nilai accuracy sebesar 0.9415, precision sebesar 0.9291, recall sebesar 0.9559, dan F1-score sebesar 0.9423. Adapun hasil confusion matrix dapat dilihat pada gambar berikut.



Gambar 4.2 Confusion Matrix Skenario 1

Berdasarkan hasil confusion matrix pada skenario 1, dari total data uji, terdapat 1094 data negatif yang berhasil diklasifikasikan dengan benar sebagai negatif (*true negative*) dan 1127 data positif yang berhasil diklasifikasikan dengan benar sebagai positif (*true positive*).

Namun demikian, masih terdapat beberapa kesalahan klasifikasi yang dilakukan oleh model. Sebanyak 86 data negatif diklasifikasikan sebagai positif (*false positive*), dan 52 data positif diklasifikasikan sebagai negatif (*false negative*).

4.2.2 Skenario Model 2

Pada skenario ini, model diuji menggunakan pembagian data sebesar 60:40, yaitu 60% sebagai data latih dan 40% sebagai data uji. Model telah melewati proses *fine-tuning* dengan batch size 16 dan learning rate sebesar 0.00003 selama 3 epoch. Hasil pengujian dirangkum pada Tabel 4.3.

Epoch	Training Loss	Validation Loss
1	0.272899	0.385409
2	0.137373	0.261055

Gambar 4.3 Hasil Training Skenario 2

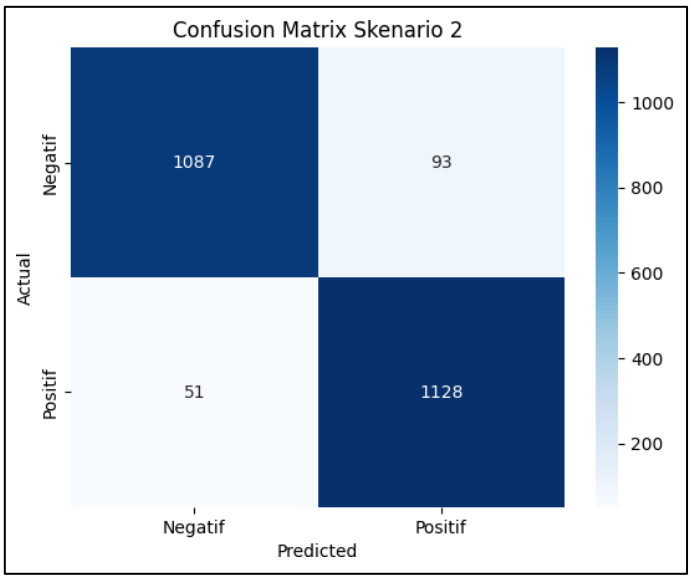
Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* dan *validation loss* yang cukup signifikan. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.272899 dan *validation loss* sebesar 0.385409. Kemudian pada *epoch* kedua nilai *training loss* mengalami penurunan menjadi 0.137373 dan *validation loss* menurun menjadi 0.261055.

Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.3.

Tabel 4.3 Hasil Evaluasi Skenario 2

Akurasi (%)	**Presisi (%)**	**Recall (%)**	**F1-Score (%)**
93.90	92.38	95.67	94.00

Berdasarkan hasil pengujian, model memperoleh nilai accuracy sebesar 0.9390, precision sebesar 0.9238, recall sebesar 0.9567, dan F1-score sebesar 0.9400. Hasil ini menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan data, baik pada kelas positif maupun negatif. Adapun hasil confusion matrix dapat dilihat pada gambar berikut.



Gambar 4.4 Confusion Matrix Skenario 2

Berdasarkan hasil confusion matrix pada skenario 2, model menunjukkan performa klasifikasi yang cukup baik dengan jumlah prediksi yang benar lebih dominan dibandingkan kesalahan prediksi.

Terdapat 1087 data negatif yang berhasil diklasifikasikan dengan benar sebagai negatif (*true negative*) dan 1128 data positif yang berhasil diklasifikasikan dengan benar sebagai positif (*true positive*). Di sisi lain, masih terdapat beberapa kesalahan klasifikasi, yaitu 93 data negatif yang diprediksi sebagai positif (*false positive*) dan 51 data positif yang diprediksi sebagai negatif (*false negative*).

4.2.3 Skenario Model 3

Pada skenario ini, model diuji menggunakan pembagian data sebesar 60:40, dengan 60% data sebagai data latih dan 40% sebagai data uji. Proses *fine-tuning* dilakukan menggunakan batch size sebesar 32 dan learning rate sebesar 0.00002 selama 2 epoch.

Epoch	Training Loss	Validation Loss
1	0.254663	0.209069
2	0.108365	0.191515

Gambar 4.5 Hasil Training Skenario 3

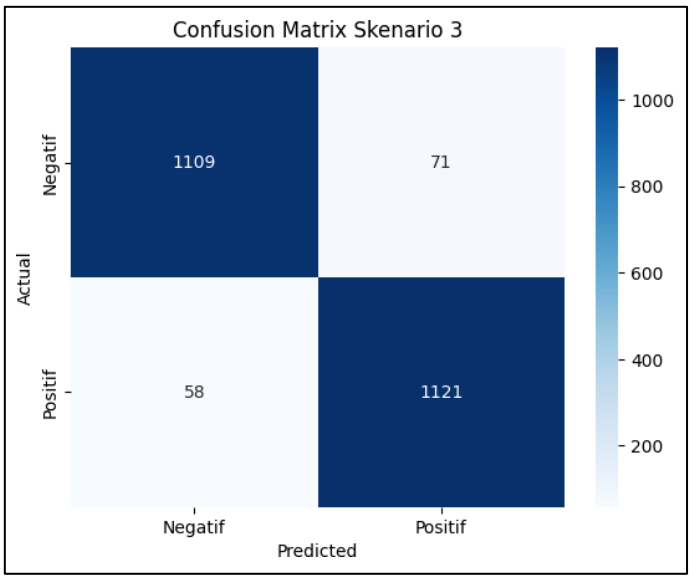
Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* dan *validation loss* yang cukup baik. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.254663 dan *validation loss* sebesar 0.209069. Kemudian pada *epoch* kedua nilai *training loss* menurun menjadi 0.108365 dan *validation loss* juga mengalami penurunan menjadi 0.191515.

Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.4.

Tabel 4.4 Hasil Evaluasi Skenario 3

Akurasi (%)	**Presisi (%)**	**Recall (%)**	**F1-Score (%)**
94.53	94.04	95.08	94.56

Berdasarkan hasil pengujian, model memperoleh nilai accuracy sebesar 0.9453, precision sebesar 0.9404, recall sebesar 0.9508, dan F1-score sebesar 0.9456. Adapun hasil confusion matrix dapat dilihat pada gambar berikut.



Gambar 4.6 Confusion Matrix Skenario 3

Berdasarkan hasil confusion matrix pada skenario 3, model menunjukkan peningkatan performa klasifikasi dibandingkan skenario sebelumnya.Terdapat 1109 data negatif yang berhasil diklasifikasikan dengan benar sebagai negatif (*true negative*) dan 1121 data positif yang berhasil diklasifikasikan dengan benar sebagai positif (*true positive*). Adapun kesalahan klasifikasi yang terjadi relatif lebih kecil, yaitu 58 data negatif yang diprediksi sebagai positif (*false positive*) dan 71 data positif yang diprediksi sebagai negatif (*false negative*).

4.2.4 Skenario Model 4

Pada skenario ini, model diuji menggunakan pembagian data sebesar 60:40, dengan 60% sebagai data latih dan 40% sebagai data uji. Proses *fine-tuning* dilakukan menggunakan batch size sebesar 32 dan learning rate sebesar 0.00003 selama 2 epoch. Hasil pengujian dirangkum pada Tabel 4.4.

Epoch	Training Loss	Validation Loss
1	0.253511	0.204317
2	0.105123	0.198227

Gambar 4.7 Hasil Training Skenario 4

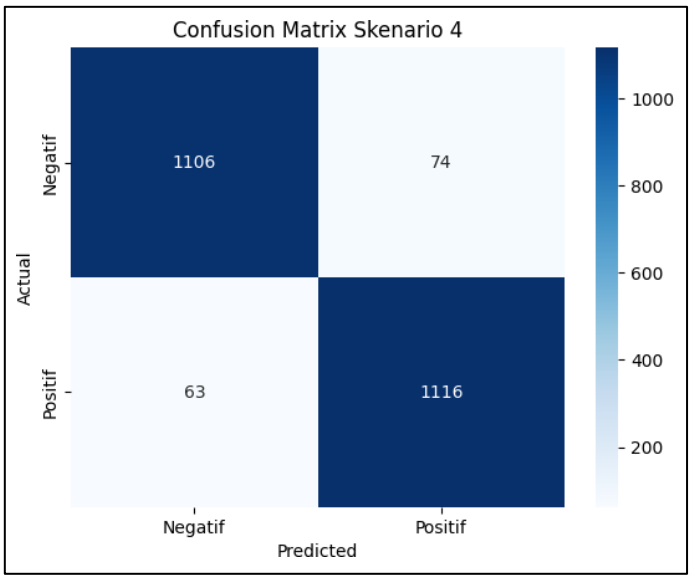
Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* dan *validation loss* yang cukup baik pada setiap *epoch*. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.253511 dan *validation loss* sebesar 0.204317. Kemudian pada *epoch* kedua nilai *training loss* mengalami penurunan menjadi 0.105123 dan *validation loss* menurun menjadi 0.198227.

Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.5.

Tabel 4.5 Hasil Evaluasi Skenario 4

Akurasi (%)	**Presisi (%)**	**Recall (%)**	**F1-Score (%)**
94.19	93.78	94.66	94.22

Berdasarkan hasil pengujian, model memperoleh nilai accuracy sebesar 0.9419, precision sebesar 0.9378, recall sebesar 0.9466, dan F1-score sebesar 0.9422. Adapun hasil confusion matrix dapat dilihat pada gambar berikut.



Gambar 4.8 Confusion Matrix Skenario 4

Berdasarkan hasil confusion matrix pada skenario 4, model menunjukkan performa klasifikasi yang cukup baik dengan dominasi prediksi yang benar pada kedua kelas. Terdapat 1106 data negatif yang berhasil diklasifikasikan dengan benar sebagai negatif (*true negative*) dan 1116 data positif yang berhasil diklasifikasikan dengan benar sebagai positif (*true positive*).

Namun demikian, masih terdapat beberapa kesalahan klasifikasi yang dilakukan oleh model. Sebanyak 74 data negatif diprediksi sebagai positif (*false positive*), dan 63 data positif diprediksi sebagai negatif (*false negative*).

4.2.5 Skenario Model 5

Pada skenario ini, model diuji menggunakan pembagian data sebesar 70:30, yaitu 70% sebagai data latih dan 30% sebagai data uji. Proses *fine-tuning* dilakukan dengan batch size sebesar 16 dan learning rate sebesar 0.00002 selama 2 epoch.

Epoch	Training Loss	Validation Loss
1	0.251757	0.206672
2	0.106536	0.255307

Gambar 4.9 Hasil Training Skenario 5

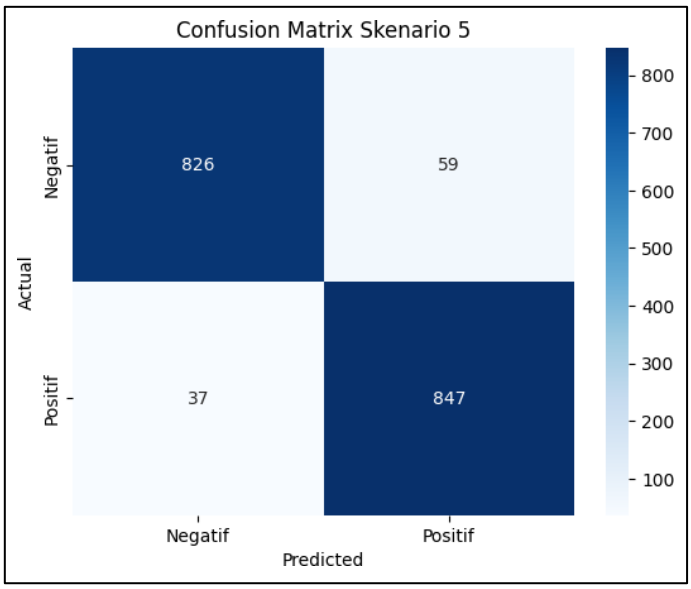
Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* pada setiap *epoch*. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.251757 dan *validation loss* sebesar 0.206672. Kemudian pada *epoch* kedua nilai *training loss* menurun menjadi 0.106536.

Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.6.

Tabel 4.6 Hasil Evaluasi Skenario 5

Akurasi (%)	**Presisi (%)**	**Recall (%)**	**F1-Score (%)**
94.57	93.49	95.81	94.64

Berdasarkan hasil pengujian, model memperoleh nilai accuracy sebesar 0.9457, precision sebesar 0.9349, recall sebesar 0.9581, dan F1-score sebesar 0.9464. Hasil ini menunjukkan bahwa model mengalami peningkatan performa dibandingkan skenario sebelumnya yang menggunakan rasio data latih lebih kecil.



Gambar 4.10 Confusion Matrix Skenario 5

Berdasarkan hasil confusion matrix pada skenario 5, model berhasil mengklasifikasikan 826 data negatif dengan benar sebagai negatif (*true negative*) dan 847 data positif dengan benar sebagai positif (*true positive*). Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali kedua kelas secara akurat. Adapun kesalahan klasifikasi yang terjadi relatif lebih kecil, yaitu sebanyak 59 data negatif yang diprediksi sebagai positif (*false positive*) dan 37 data positif yang diprediksi sebagai negatif (*false negative*).

4.2.6 Skenario Model 6

Pada skenario ini, model diuji menggunakan pembagian data sebesar 70:30, dengan 70% data sebagai data latih dan 30% sebagai data uji. Proses pelatihan dilakukan menggunakan batch size sebesar 16 dan learning rate sebesar 0.00003 selama 2 epoch.

Epoch	Training Loss	Validation Loss
1	0.272065	0.227142
2	0.086096	0.270437

Gambar 4.11 Hasil Training Skenario 6

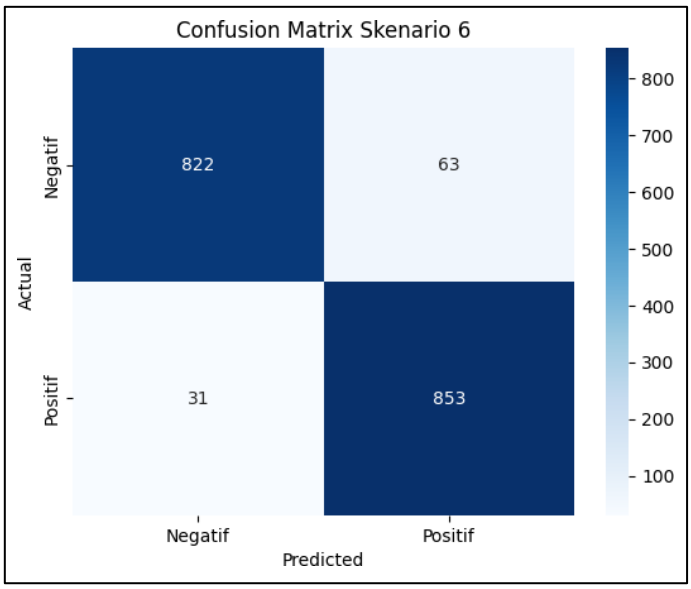
Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* yang cukup signifikan. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.272065 dan *validation loss* sebesar 0.227142. Kemudian pada *epoch* kedua nilai *training loss* mengalami penurunan menjadi 0.086096.

Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.7.

Tabel 4.7 Hasil Evaluasi Skenario 6

Akurasi (%)	**Presisi (%)**	**Recall (%)**	**F1-Score (%)**
94.69	93.12	96.49	94.78

Berdasarkan hasil pengujian, model memperoleh nilai accuracy sebesar 0.9469, precision sebesar 0.9312, recall sebesar 0.9649, dan F1-score sebesar 0.9478. Hasil ini menunjukkan adanya peningkatan dibanding skenario 5.



Gambar 4.12 Confusion Matrix Skenario 6

Berdasarkan hasil confusion matrix pada skenario 6, Model berhasil mengklasifikasikan 822 data negatif dengan benar sebagai negatif (*true negative*) dan 853 data positif dengan benar sebagai positif (*true positive*). Adapun kesalahan klasifikasi yang terjadi terdiri dari 63 data negatif yang diprediksi sebagai positif (*false positive*) dan 31 data positif yang diprediksi sebagai negatif (*false negative*).

4.2.7 Skenario Model 7

Pada skenario ini, model diuji menggunakan pembagian data sebesar 70:30, dengan 70% data sebagai data latih dan 30% sebagai data uji. Proses pelatihan dilakukan menggunakan batch size sebesar 32 dan learning rate sebesar 0.00002 selama 2 epoch. Hasil pengujian dirangkum pada Tabel 4.7.

Epoch	Training Loss	Validation Loss
1	0.279804	0.206969
2	0.092441	0.218896

Gambar 4.13 Hasil Training Skenario 7

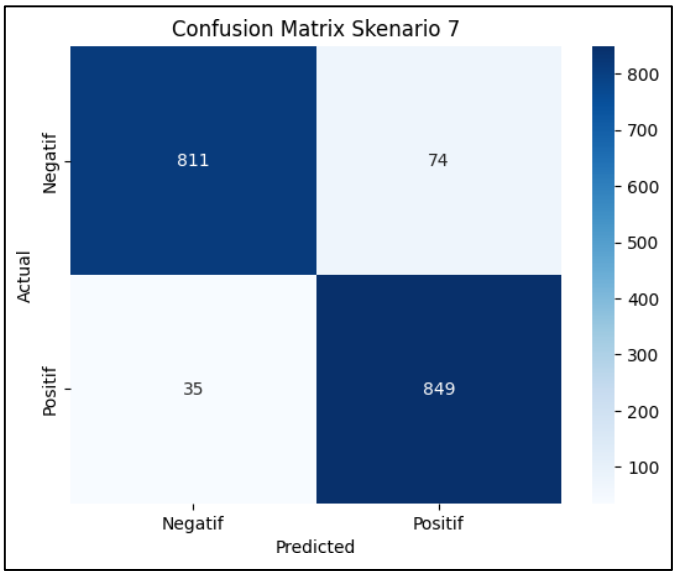
Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* yang cukup signifikan. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.279804 dan *validation loss* sebesar 0.206969. Kemudian pada *epoch* kedua nilai *training loss* mengalami penurunan menjadi 0.092441 dan *validation loss* meningkat sedikit menjadi 0.218896.

Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.8.

Tabel 4.8 Hasil Evaluasi Skenario 7

Akurasi (%)	**Presisi (%)**	**Recall (%)**	**F1-Score (%)**
93.84	91.98	96.04	93.97

Berdasarkan hasil pengujian, model memperoleh nilai accuracy sebesar 0.9384, precision sebesar 0.9198, recall sebesar 0.9604, dan F1-score sebesar 0.9397. Adapun hasil confusion matrix dapat dilihat pada gambar berikut.



Gambar 4.14 Confusion Matrix Skenario 7

Berdasarkan hasil confusion matrix pada skenario 7, model berhasil mengklasifikasikan 811 data negatif dengan benar sebagai negatif (*true negative*) dan 849 data positif dengan benar sebagai positif (*true positive*). Namun, terdapat kesalahan klasifikasi yang sedikit lebih tinggi, yaitu 74 data negatif yang diprediksi sebagai positif (*false positive*) dan 35 data positif yang diprediksi sebagai negatif (*false negative*).

4.2.8 Skenario Model 8

Pada skenario ini, model diuji menggunakan pembagian data sebesar 70:30, dengan 70% data sebagai data latih dan 30% sebagai data uji. Proses pelatihan dilakukan menggunakan batch size sebesar 32 dan learning rate sebesar 0.00003 selama 2 epoch.

Epoch	Training Loss	Validation Loss
1	0.276488	0.203819
2	0.078130	0.224486

Gambar 4.15 Hasil Training Skenario 8

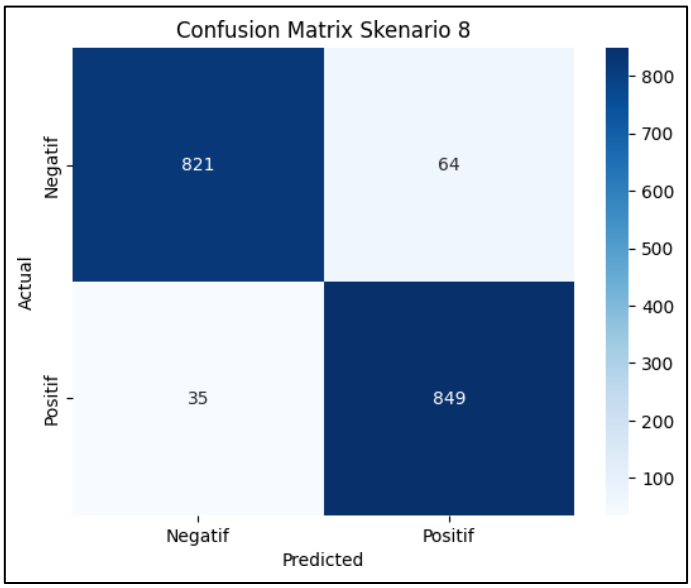
Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* yang cukup signifikan pada setiap *epoch*. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.276488 dan *validation loss* sebesar 0.203819. Kemudian pada *epoch* kedua nilai *training loss* menurun menjadi 0.078130, sedangkan *validation loss* mengalami sedikit peningkatan menjadi 0.224486.

Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.9.

Tabel 4.9 Hasil Evaluasi Skenario 8

Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
94.40	92.99	96.04	94.49

Berdasarkan hasil pengujian, model memperoleh nilai accuracy sebesar 0.9440, precision sebesar 0.9299, recall sebesar 0.9604, dan F1-score sebesar 0.9449. Hasil ini menunjukkan adanya peningkatan performa dibandingkan skenario 7, terutama pada nilai accuracy, precision, dan F1-score.



Gambar 4.16 Confusion Matrix Skenario 8

Berdasarkan hasil confusion matrix pada skenario 8, model berhasil mengklasifikasikan 821 data negatif dengan benar sebagai negatif (*true negative*) dan 849 data positif dengan benar sebagai positif (*true positive*). Adapun kesalahan klasifikasi yang terjadi terdiri dari 64 data negatif yang diprediksi sebagai positif (*false positive*) dan 35 data positif yang diprediksi sebagai negatif (*false negative*).

4.2.9 Skenario Model 9

Pada skenario ini, model diuji menggunakan pembagian data sebesar 80:20, dengan 80% data sebagai data latih dan 20% sebagai data uji. Proses pelatihan dilakukan menggunakan batch size sebesar 16 dan learning rate sebesar 0.00002 selama 2 epoch. Hasil pengujian dirangkum pada Tabel 4.9.

Epoch	Training Loss	Validation Loss
1	0.203799	0.171127
2	0.079896	0.206800

Gambar 4.17 Hasil Training Skenario 9

Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* yang cukup baik pada setiap *epoch*. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.203799 dan *validation loss* sebesar 0.171127. Kemudian pada *epoch* kedua nilai *training loss* mengalami penurunan menjadi 0.079896, sedangkan *validation loss* meningkat menjadi 0.206800.

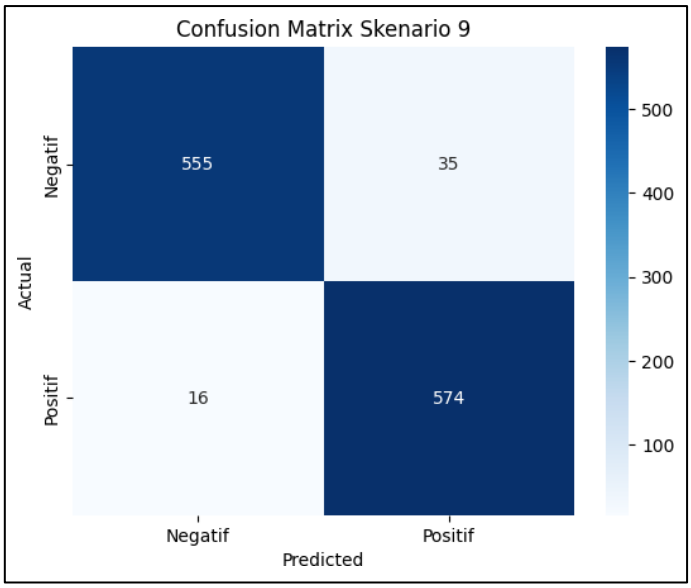
Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.10.

Tabel 4.10 Hasil Evaluasi Skenario 9

Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
95.68	94.25	97.29	95.75

Berdasarkan hasil pengujian, model memperoleh nilai accuracy sebesar 0.9568, precision sebesar 0.9425, recall sebesar 0.9729, dan F1-score sebesar

0.9575. Peningkatan performa ini disebabkan oleh penggunaan proporsi data latih yang lebih besar.



Gambar 4.18 Confusion Matrix Skenario 9

Berdasarkan hasil confusion matrix pada skenario 9, model berhasil mengklasifikasikan 555 data negatif dengan benar sebagai negatif (*true negative*) dan 574 data positif dengan benar sebagai positif (*true positive*).

Adapun kesalahan klasifikasi yang terjadi terdiri dari 35 data negatif yang diprediksi sebagai positif (*false positive*) dan 16 data positif yang diprediksi sebagai negatif (*false negative*). Jumlah kesalahan ini tergolong kecil, terutama pada false negative yang sangat rendah.

4.2.10 Skenario Model 10

Pada skenario ini, model diuji menggunakan pembagian data sebesar 80:20, dengan 80% data sebagai data latih dan 20% sebagai data uji. Proses pelatihan dilakukan menggunakan batch size sebesar 16 dan learning rate sebesar 0.00003 selama 2 epoch.

Epoch	Training Loss	Validation Loss
1	0.205498	0.145210
2	0.085270	0.221321

Gambar 4.19 Hasil Training Skenario 10

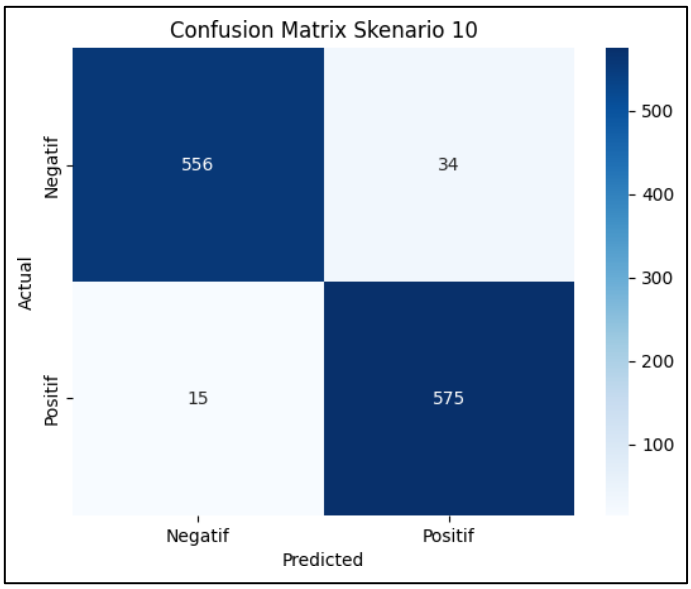
Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* yang cukup signifikan pada setiap *epoch*. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.205498 dan *validation loss* sebesar 0.145210. Kemudian pada *epoch* kedua nilai *training loss* mengalami penurunan menjadi 0.085270, sedangkan *validation loss* meningkat menjadi 0.221321.

Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.11.

Tabel 4.11 Hasil Evaluasi Skenario 10

Akurasi (%)	**Presisi (%)**	**Recall (%)**	**F1-Score (%)**
95.85	94.42	97.46	95.91

Berdasarkan hasil pengujian, model memperoleh nilai accuracy sebesar 0.9585, precision sebesar 0.9442, recall sebesar 0.9746, dan F1-score sebesar 0.9591. Adapun hasil confusion matrix dapat dilihat pada gambar berikut.



Gambar 4.20 Confusion Matrix Skenario 10

Berdasarkan hasil confusion matrix pada skenario 10, model berhasil mengklasifikasikan 556 data negatif dengan benar sebagai negatif (*true negative*) dan 575 data positif dengan benar sebagai positif (*true positive*). Adapun kesalahan klasifikasi yang terjadi terdiri dari 34 data negatif yang diprediksi sebagai positif (*false positive*) dan 15 data positif yang diprediksi sebagai negatif (*false negative*).

4.2.11 Skenario Model 11

Pada skenario ini, model diuji menggunakan pembagian data sebesar 80:20, dengan 80% data sebagai data latih dan 20% sebagai data uji. Berbeda dengan skenario sebelumnya, pada skenario ini digunakan batch size sebesar 32 dan learning rate sebesar 0.00002 selama 2 epoch.

Epoch	Training Loss	Validation Loss
1	0.293052	0.158600
2	0.084592	0.157184

Gambar 4.21 Hasil Training Skenario 11

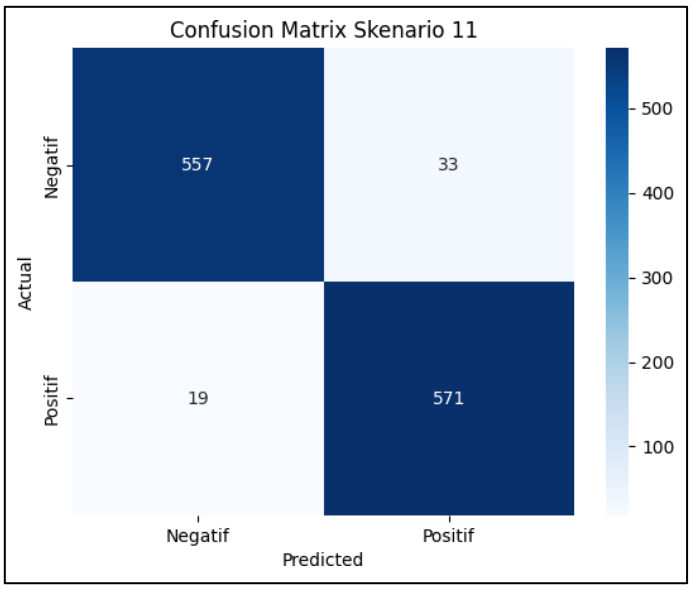
Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* dan *validation loss* yang cukup baik pada setiap *epoch*. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.293052 dan *validation loss* sebesar 0.158600. Kemudian pada *epoch* kedua nilai *training loss* mengalami penurunan menjadi 0.084592 dan *validation loss* juga menurun menjadi 0.157184.

Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.12.

Tabel 4.12 Hasil Evaluasi Skenario 11

Akurasi (%)	**Presisi (%)**	**Recall (%)**	**F1-Score (%)**
95.59	94.54	96.78	95.64

Berdasarkan hasil pengujian, model memperoleh accuracy sebesar 0.9559, precision sebesar 0.9454, recall sebesar 0.9678, dan F1-score sebesar 0.9564. Secara umum, performa model masih tergolong tinggi dan stabil di semua metrik evaluasi. Adapun hasil confusion matrix dapat dilihat pada gambar berikut.



Gambar 4.22 Confusion Matrix Skenario 11

Berdasarkan hasil confusion matrix pada skenario 11, Model berhasil mengklasifikasikan 557 data negatif dengan benar sebagai negatif (*true negative*) dan 571 data positif dengan benar sebagai positif (*true positive*).

Adapun kesalahan klasifikasi yang terjadi terdiri dari 33 data negatif yang diprediksi sebagai positif (*false positive*) dan 19 data positif yang diprediksi sebagai negatif (*false negative*). Jumlah kesalahan ini tergolong kecil jika dibandingkan dengan jumlah prediksi yang benar, sehingga dapat dikatakan bahwa model memiliki tingkat akurasi yang tinggi.

4.2.12 Skenario Model 12

Pada skenario ini, model diuji menggunakan pembagian data sebesar 80:20, dengan 80% data sebagai data latih dan 20% sebagai data uji. Proses pelatihan dilakukan menggunakan batch size sebesar 32 dan learning rate sebesar 0.00003 selama 2 epoch.

Epoch	Training Loss	Validation Loss
1	0.284018	0.165855
2	0.072335	0.165793

Gambar 4.23 Hasil Training Skenario 12

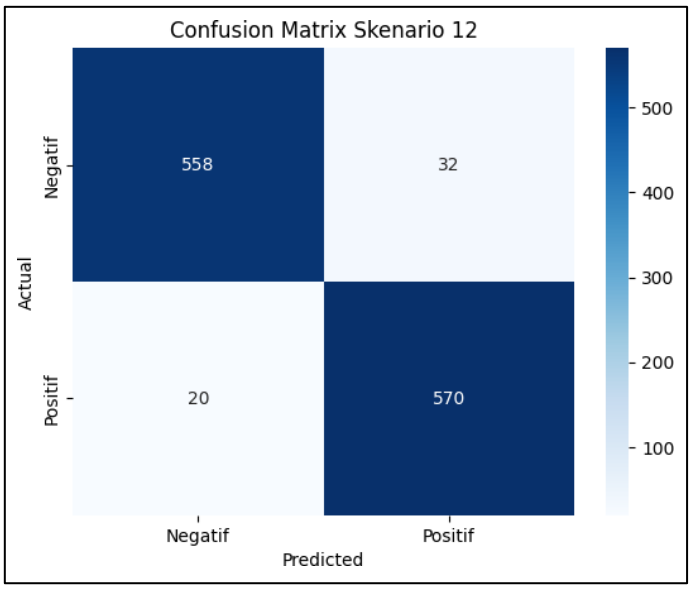
Selama proses pelatihan berlangsung, model menunjukkan penurunan nilai *training loss* yang cukup signifikan pada setiap *epoch*. Pada *epoch* pertama diperoleh nilai *training loss* sebesar 0.284018 dan *validation loss* sebesar 0.165855. Kemudian pada *epoch* kedua nilai *training loss* mengalami penurunan menjadi 0.072335, sedangkan *validation loss* juga mengalami sedikit penurunan menjadi 0.165793.

Setelah proses pelatihan, model kemudian dievaluasi menggunakan data uji untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Hasil evaluasi ditunjukkan pada Tabel 4.13.

Tabel 4.13 Hasil Evaluasi Skenario 12

Akurasi (%)	**Presisi (%)**	**Recall (%)**	**F1-Score (%)**
95.59	94.68	96.61	95.64

Berdasarkan hasil pengujian, model memperoleh nilai accuracy sebesar 0.9559, precision sebesar 0.9468, recall sebesar 0.9661, dan F1-score sebesar 0.9564. Jika dibandingkan dengan skenario 11, tidak terdapat perbedaan yang signifikan pada nilai accuracy dan F1-score, namun terdapat sedikit peningkatan pada nilai precision. Adapun hasil confusion matrix dapat dilihat pada gambar berikut.



Gambar 4.24 Confusion Matrix Skenario 12

Berdasarkan hasil confusion matrix pada skenario 12, model berhasil mengklasifikasikan 558 data negatif dengan benar sebagai negatif (*true negative*) dan 570 data positif dengan benar sebagai positif (*true positive*).

Adapun kesalahan klasifikasi yang terjadi terdiri dari 32 data negatif yang diprediksi sebagai positif (*false positive*) dan 20 data positif yang diprediksi sebagai negatif (*false negative*).

4.3 Pembahasan

Berdasarkan hasil pengujian yang telah dilakukan pada 12 skenario, diperoleh performa model yang secara umum menunjukkan hasil yang baik dan konsisten. Evaluasi dilakukan menggunakan empat metrik utama yaitu accuracy, precision, recall, dan F1-score untuk mengukur kinerja model dalam melakukan klasifikasi sentimen. Untuk memberikan gambaran yang lebih jelas mengenai hasil pengujian, berikut disajikan tabel perbandingan performa model.

Tabel 4.14 Hasil Performa Seluruh Skenario

Skenario	Rasio	Batch Size	Learning Rate	Accuracy	Precision	Recall	F1-Score
Skenario 1	60:40	16	0.00002	0.9415	0.9291	0.9559	0.9423
Skenario 2	60:40	16	0.00003	0.9390	0.9238	0.9567	0.9400
Skenario 3	60:40	32	0.00002	0.9453	0.9404	0.9508	0.9456
Skenario 4	60:40	32	0.00003	0.9419	0.9378	0.9466	0.9422
Skenario 5	70:30	16	0.00002	0.9457	0.9349	0.9581	0.9464
Skenario 6	70:30	16	0.00003	0.9469	0.9312	0.9649	0.9478
Skenario 7	70:30	32	0.00002	0.9384	0.9198	0.9604	0.9397
Skenario 8	70:30	32	0.00003	0.9440	0.9299	0.9604	0.9449
Skenario 9	80:20	16	0.00002	0.9568	0.9425	0.9729	0.9575
Skenario 10	80:20	16	0.00003	0.9585	0.9442	0.9746	0.9591
Skenario 11	80:20	32	0.00002	0.9559	0.9454	0.9678	0.9564
Skenario 12	80:20	32	0.00003	0.9559	0.9468	0.9661	0.9564

Berdasarkan tabel hasil evaluasi seluruh skenario tersebut, dapat dilihat bahwa model telah diuji menggunakan berbagai kombinasi parameter yang meliputi rasio pembagian data, batch size, dan learning rate. Setiap kombinasi tersebut menghasilkan performa yang berbeda-beda, namun secara umum menunjukkan hasil yang cukup konsisten.

Jika dilihat dari keseluruhan hasil, model mampu mencapai performa yang tinggi pada semua skenario, dengan nilai accuracy yang berada di atas 93%. Hal ini menunjukkan bahwa model yang digunakan sudah mampu mempelajari pola dari data dengan baik dan dapat melakukan klasifikasi sentimen secara akurat.

Pada rasio 60:40, performa model sudah tergolong baik dengan nilai accuracy berada di kisaran 0.9390 hingga 0.9453. Pada rasio ini, data latih yang digunakan masih relatif lebih sedikit dibandingkan rasio lainnya, sehingga model memiliki keterbatasan dalam mempelajari pola secara lebih mendalam. Meskipun demikian, hasil yang diperoleh tetap stabil dan menunjukkan bahwa model sudah cukup mampu melakukan klasifikasi dengan baik.

Selanjutnya pada rasio 70:30, performa model mengalami sedikit peningkatan dengan nilai accuracy berada di kisaran 0.9384 hingga 0.9469. Dengan jumlah data latih yang lebih banyak dibandingkan rasio 60:40, model dapat belajar lebih baik sehingga menghasilkan performa yang sedikit lebih tinggi. Selain itu, nilai recall pada rasio ini juga cenderung tinggi, yang menunjukkan bahwa model semakin baik dalam mendeteksi data positif.

Pada rasio 80:20, performa model menunjukkan hasil terbaik dibandingkan rasio lainnya, dengan nilai accuracy mencapai 0.9568 hingga 0.9585. Hal ini menunjukkan bahwa semakin banyak data latih yang digunakan, maka kemampuan model dalam mengenali pola menjadi semakin baik. Dengan data latih yang lebih besar, model dapat melakukan generalisasi dengan lebih optimal sehingga menghasilkan performa yang lebih tinggi pada data uji.

Hasil pengujian menunjukkan bahwa setiap kombinasi parameter yang digunakan menghasilkan performa yang berbeda-beda. Perbedaan tersebut terlihat dari nilai accuracy, precision, recall, dan F1-score yang diperoleh pada masing-masing skenario. Meskipun seluruh skenario mampu memberikan performa yang baik, kombinasi parameter pada skenario 10 menghasilkan performa yang paling

optimal dibandingkan skenario lainnya. Hal ini menunjukkan bahwa pemilihan parameter yang tepat memiliki pengaruh penting terhadap keberhasilan model dalam melakukan klasifikasi sentimen. Temuan tersebut sejalan dengan firman Allah SWT dalam QS. Al-Qamar ayat 49:

إِنَّا كُلَّ شَيْءٍ خَلَقْنَٰهُ بِقَدَرٍ

“ Sesungguhnya Kami menciptakan segala sesuatu sesuai dengan ukuran.”
(QS Al-Qamar : 49)

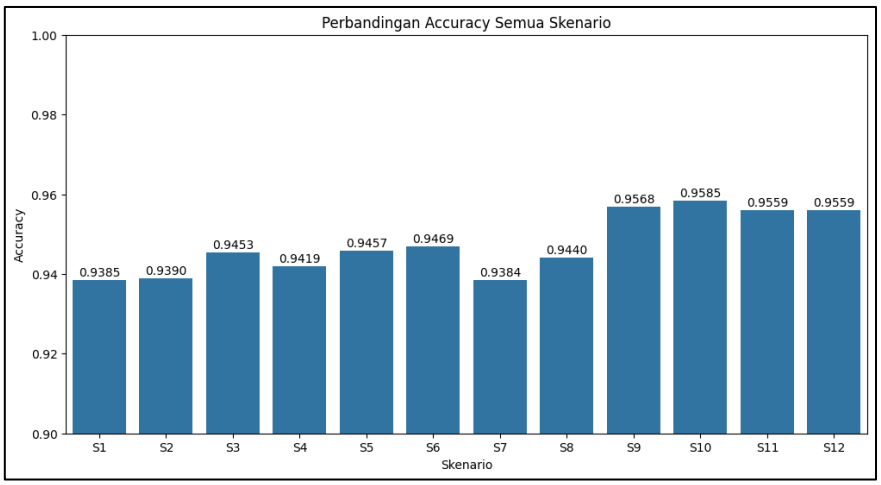
Ayat tersebut menjelaskan bahwa segala sesuatu yang terjadi telah ditetapkan oleh Allah dengan ukuran, sistem, dan ketentuan tertentu. Dalam konteks penelitian ini, konsep tersebut dapat dikaitkan dengan proses pelatihan model IndoBERT yang dilakukan berdasarkan parameter dan aturan yang telah ditentukan, seperti rasio pembagian data, batch size, dan learning rate. Setiap parameter memberikan pengaruh yang berbeda terhadap proses pembelajaran model sehingga menghasilkan performa yang berbeda.

Hasil penelitian menunjukkan bahwa kombinasi parameter pada skenario 10, yaitu rasio data 80:20, batch size 16, dan learning rate 0,00003 mampu menghasilkan performa terbaik dibandingkan skenario lainnya. Hal ini menunjukkan bahwa untuk memperoleh hasil yang optimal diperlukan pemilihan parameter yang sesuai dengan karakteristik data dan model yang digunakan.

Dengan demikian, hasil penelitian ini mencerminkan bahwa suatu proses dapat menghasilkan keluaran yang optimal apabila dijalankan berdasarkan sistem, ukuran, dan ketentuan yang tepat, sebagaimana dijelaskan dalam QS. Al-Qamar ayat 49.

4.3.1 Akurasi

Visualisasi akurasi dilakukan untuk melihat tingkat ketepatan model IndoBERT dalam melakukan klasifikasi sentimen pada seluruh skenario pengujian. Nilai akurasi menunjukkan seberapa besar kemampuan model dalam memprediksi data sentimen secara benar dibandingkan keseluruhan data uji.



Gambar 4.24 Grafik Perbandingan Akurasi Seluruh Skenario

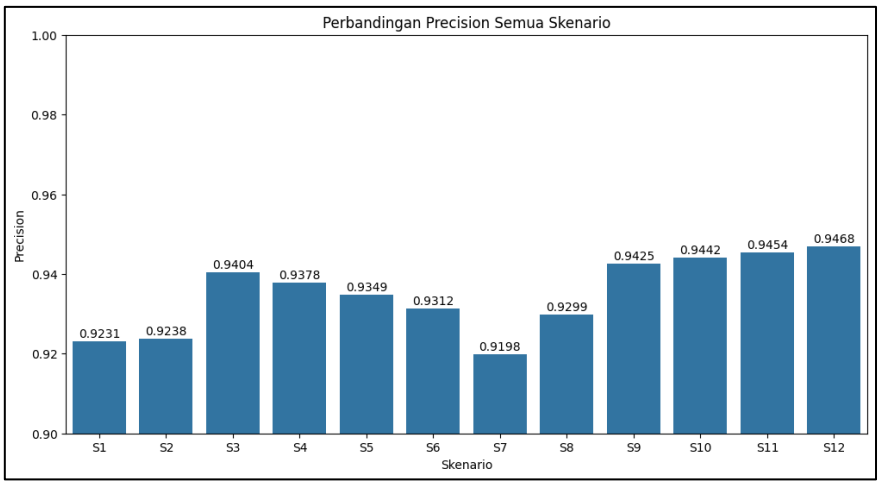
Berdasarkan grafik perbandingan akurasi seluruh skenario, terlihat bahwa seluruh model menghasilkan performa yang cukup baik dengan nilai akurasi di atas 93%.Nilai akurasi tertinggi diperoleh pada skenario 10 dengan nilai sebesar 0.9585. Hasil tersebut menunjukkan bahwa kombinasi rasio data 80:20, *batch size* sebesar 16, dan *learning rate* sebesar 0.00003 mampu menghasilkan performa klasifikasi yang paling optimal dibandingkan skenario lainnya.

Sementara itu, nilai akurasi terendah diperoleh pada skenario 7 dengan nilai sebesar 0.9384. Meskipun demikian, nilai tersebut masih tergolong cukup baik dalam proses klasifikasi sentimen. Secara keseluruhan, hasil visualisasi menunjukkan bahwa penggunaan rasio data 80:20 cenderung menghasilkan nilai

akurasi yang lebih tinggi dibandingkan rasio data lainnya karena model memperoleh jumlah data latih yang lebih banyak selama proses pembelajaran.

4.3.2 Presisi

Visualisasi *precision* dilakukan untuk melihat tingkat ketepatan model IndoBERT dalam memprediksi data sentimen positif pada seluruh skenario pengujian. Nilai *precision* menunjukkan seberapa banyak prediksi positif yang benar dibandingkan seluruh data yang diprediksi positif oleh model.

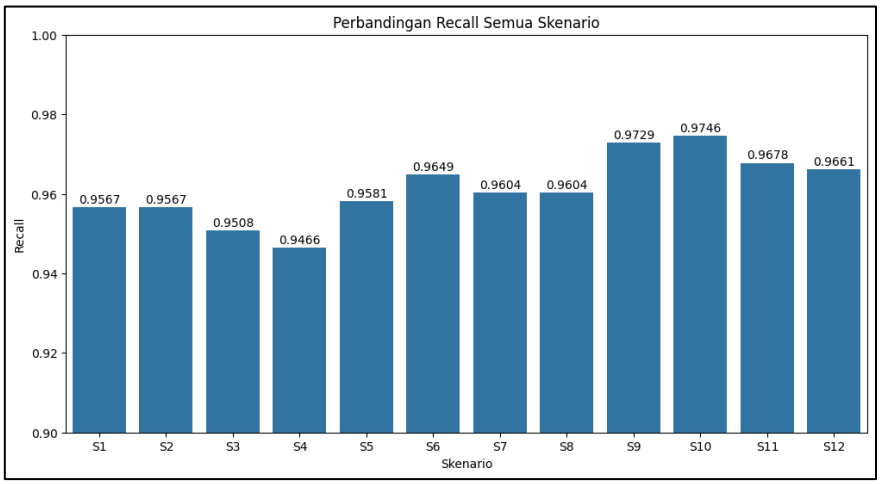


Gambar 4.25 Grafik Perbandingan Presisi Seluruh Skenario

Berdasarkan grafik perbandingan *precision* seluruh skenario, nilai *precision* tertinggi diperoleh pada skenario 12 dengan nilai sebesar 0.9468. Hasil tersebut menunjukkan bahwa kombinasi rasio data 80:20, *batch size* sebesar 32, dan *learning rate* sebesar 0.00003 mampu menghasilkan tingkat ketepatan prediksi positif yang paling baik dibandingkan skenario lainnya. Tingginya nilai *precision* menunjukkan bahwa sebagian besar data yang diprediksi sebagai sentimen positif oleh model memang benar termasuk ke dalam kelas positif. Sementara itu, nilai *precision* terendah diperoleh pada skenario 7 dengan nilai sebesar 0.9198.

4.3.3 Recall

Visualisasi *recall* dilakukan untuk melihat kemampuan model IndoBERT dalam mengidentifikasi seluruh data sentimen positif pada setiap skenario pengujian. Nilai *recall* menunjukkan seberapa banyak data positif yang berhasil diprediksi dengan benar oleh model dibandingkan seluruh data yang sebenarnya termasuk dalam kelas positif.



Gambar 4.26 Grafik Perbandingan Recall Seluruh Skenario

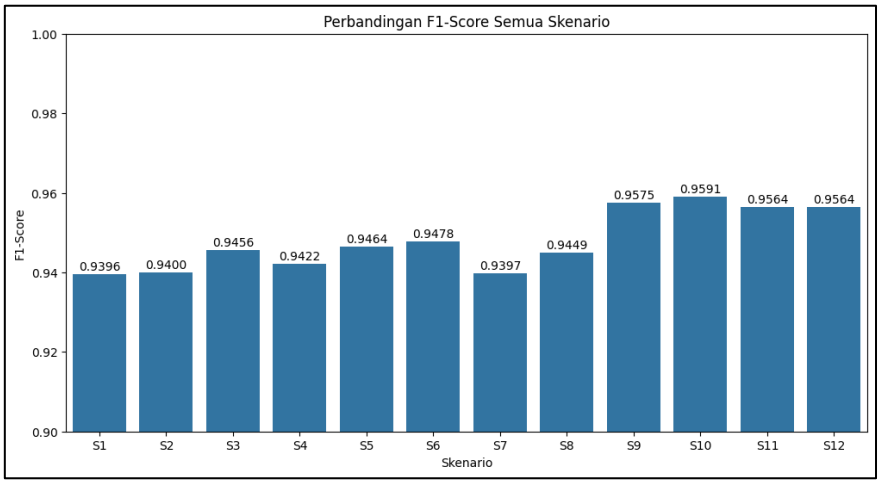
Berdasarkan grafik perbandingan *recall* seluruh skenario, terlihat bahwa seluruh model menghasilkan nilai *recall* yang sangat tinggi dengan rentang nilai antara 0.9466 hingga 0.9746. Nilai *recall* tertinggi diperoleh pada skenario 10 dengan nilai sebesar 0.9746.

Hasil tersebut menunjukkan bahwa kombinasi parameter pada skenario tersebut mampu menghasilkan kemampuan terbaik dalam mendeteksi data sentimen positif. Tingginya nilai *recall* menunjukkan bahwa sebagian besar data positif berhasil dikenali dan diklasifikasikan dengan benar oleh model.

Sementara itu, nilai *recall* terendah diperoleh pada skenario 4 dengan nilai sebesar 0.9466. Meskipun menjadi nilai terendah dibandingkan skenario lainnya, performa tersebut masih tergolong sangat baik karena nilai *recall* tetap berada di atas 0.94.

4.3.4 F1-Score

Visualisasi *F1-Score* dilakukan untuk mengevaluasi keseimbangan antara nilai *precision* dan *recall* yang dihasilkan oleh model IndoBERT pada seluruh skenario pengujian. Nilai *F1-Score* merupakan rata-rata harmonis dari *precision* dan *recall*, sehingga dapat memberikan gambaran performa model secara lebih menyeluruh dalam proses klasifikasi sentimen.



Gambar 4.27 Grafik Perbandingan F1-Score Seluruh Skenario

Berdasarkan grafik perbandingan *F1-Score* seluruh skenario, terlihat bahwa seluruh model menghasilkan nilai *F1-Score* yang tinggi dengan rentang nilai antara 0.9396 hingga 0.9591. Hasil tersebut menunjukkan bahwa model IndoBERT mampu menjaga keseimbangan yang baik antara ketepatan prediksi positif (*precision*) dan kemampuan mendeteksi seluruh data positif (*recall*).

Nilai *F1-Score* tertinggi diperoleh pada skenario 10 dengan nilai sebesar 0.9591. Hasil ini menunjukkan bahwa kombinasi parameter pada skenario tersebut mampu menghasilkan performa klasifikasi yang paling optimal dibandingkan skenario lainnya. Nilai *F1-Score* yang tinggi mengindikasikan bahwa model tidak hanya memiliki tingkat ketepatan prediksi yang baik, tetapi juga mampu mengenali sebagian besar data positif dengan benar. Sementara itu, nilai *F1-Score* terendah diperoleh pada skenario 1 dengan nilai sebesar 0.9396.

4.3.5 Pengaruh Rasio Pembagian Data

Berdasarkan hasil pengujian yang telah dilakukan, terlihat bahwa rasio pembagian data memberikan pengaruh terhadap performa model yang dihasilkan. Secara umum, semakin besar proporsi data latih yang digunakan, maka performa model cenderung mengalami peningkatan. Untuk melihat perbandingan secara lebih jelas, disajikan pada tabel berikut.

Tabel 4.15 Rata-Rata Performa Berdasarkan Rasio Data

Rasio Data	Rata-rata Accuracy	Rata-rata Precision	Rata-rata Recall	Rata-rata F1-Score
60:40	0.9419	0.9328	0.9525	0.9425
70:30	0.9438	0.9290	0.9609	0.9447
80:20	0.9568	0.9447	0.9704	0.9573

Pada rasio 60:40, model dilatih menggunakan 60% data dan diuji menggunakan 40% data. Hasil pengujian menunjukkan bahwa model sudah mampu mencapai performa yang cukup baik dengan nilai accuracy di kisaran 0.9390 hingga 0.9453. Namun, dibandingkan dengan rasio lainnya, performa pada rasio ini masih relatif lebih rendah. Hal ini disebabkan karena jumlah data latih yang digunakan

masih terbatas, sehingga model belum sepenuhnya mampu menangkap pola secara optimal.

Selanjutnya, pada rasio 70:30, terjadi peningkatan performa model dengan nilai accuracy yang berada pada kisaran 0.9384 hingga 0.9469. Dengan jumlah data latih yang lebih banyak dibandingkan rasio 60:40, model memiliki kesempatan yang lebih besar untuk mempelajari karakteristik data. Hal ini berdampak pada peningkatan kemampuan model dalam melakukan klasifikasi, terutama dalam mendeteksi data positif yang ditunjukkan oleh nilai recall yang cenderung tinggi.

Pada rasio 80:20, model menunjukkan performa terbaik dengan nilai accuracy mencapai 0.9568 hingga 0.9585. Hal ini menunjukkan bahwa penggunaan data latih yang lebih besar memberikan dampak yang signifikan terhadap peningkatan performa model. Dengan lebih banyak data yang dipelajari, model dapat mengenali pola dengan lebih baik dan menghasilkan prediksi yang lebih akurat pada data uji.

4.3.6 Pengaruh Batch Size

Batch size merupakan salah satu parameter penting dalam proses pelatihan model, yang menentukan jumlah data yang diproses dalam satu iterasi sebelum dilakukan pembaruan bobot. Pada penelitian ini, digunakan dua variasi batch size yaitu 16 dan 32.

Berdasarkan hasil pengujian yang telah dilakukan pada seluruh skenario, terlihat bahwa batch size memberikan pengaruh terhadap performa model, meskipun tidak terlalu signifikan dibandingkan rasio pembagian data. Untuk melihat perbandingan secara lebih jelas, disajikan pada tabel berikut.

Tabel 4.16 Rata-Rata Performa Berdasarkan Batch Size

Batch Size	Rata-rata Accuracy	Rata-rata Precision	Rata-rata Recall	Rata-rata F1-Score
16	0.9481	0.9360	0.9639	0.9492
32	0.9469	0.9367	0.9587	0.9479

Berdasarkan tabel tersebut, dapat dilihat bahwa batch size 16 menghasilkan performa yang sedikit lebih baik dibandingkan batch size 32, terutama pada nilai accuracy, recall, dan F1-score. Hal ini menunjukkan bahwa penggunaan batch size yang lebih kecil mampu memberikan hasil yang lebih optimal dalam penelitian ini.

Batch size yang lebih kecil memungkinkan model untuk melakukan pembaruan bobot lebih sering dalam satu epoch, sehingga model menjadi lebih responsif terhadap pola-pola kecil dalam data. Hal ini berdampak pada meningkatnya kemampuan model dalam mendeteksi data positif, yang terlihat dari nilai recall yang lebih tinggi.

Sementara itu, batch size 32 menunjukkan performa yang relatif stabil, namun sedikit lebih rendah dibandingkan batch size 16. Hal ini disebabkan karena batch size yang lebih besar cenderung membuat proses pembelajaran menjadi lebih halus, tetapi kurang sensitif terhadap variasi data yang lebih kecil.

Perbedaan performa antara kedua batch size tidak terlalu besar, yang menunjukkan bahwa model masih cukup stabil terhadap perubahan parameter ini. Dengan kata lain, baik batch size 16 maupun 32 masih berada dalam rentang yang optimal untuk digunakan.

4.3.7 Pengaruh Learning Rate

Pada penelitian ini digunakan dua nilai learning rate, yaitu 0.00002 dan 0.00003, untuk melihat pengaruhnya terhadap performa klasifikasi sentimen. Berdasarkan hasil pengujian, penggunaan learning rate 0.00003 cenderung memberikan hasil yang sedikit lebih baik dibandingkan 0.00002.

Learning rate yang lebih besar memungkinkan model untuk belajar lebih cepat dalam jumlah epoch yang terbatas. Hal ini terlihat pada beberapa skenario seperti skenario 8, 10, dan 12, di mana terdapat peningkatan pada nilai precision maupun F1-score.

Tabel 4.17 Rata-Rata Performa Berdasarkan Learning Rate

Learning Rate	Accuracy	Precision	Recall	F1-Score
0.00002	0.9473	0.9354	0.9609	0.9479
0.00003	0.9477	0.9356	0.9616	0.9484

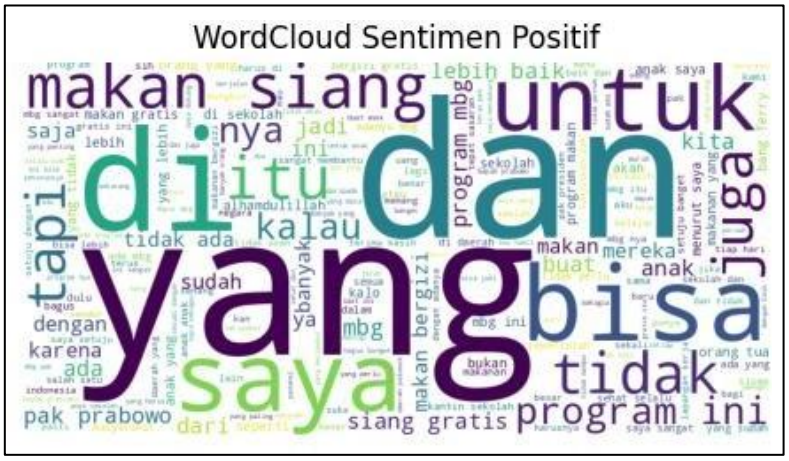
Pada rasio data 60:40 dengan batch size 16, penggunaan learning rate 0.00002 menghasilkan accuracy sebesar 0.9415 dan F1-score sebesar 0.9423. Sedangkan pada learning rate 0.00003, accuracy sedikit menurun menjadi 0.9390 dan F1-score menjadi 0.9400. Hal ini menunjukkan bahwa pada kombinasi parameter tersebut, learning rate yang lebih kecil memberikan performa yang lebih stabil.

Namun, pada rasio data 70:30 dengan batch size 16, penggunaan learning rate 0.00003 justru memberikan hasil yang lebih baik dibandingkan 0.00002. Accuracy meningkat dari 0.9457 menjadi 0.9469 dan F1-score meningkat dari 0.9464 menjadi 0.9478. Kondisi serupa juga terlihat pada rasio data 80:20, dimana

learning rate 0.00003 menghasilkan performa terbaik penelitian pada skenario 10 dengan accuracy sebesar 0.9585 dan F1-score sebesar 0.9591.

4.4 WordCloud Sentimen

Visualisasi wordcloud digunakan untuk memberikan gambaran mengenai kata-kata yang paling sering muncul pada masing-masing kelas sentimen, yaitu sentimen positif dan sentimen negatif. Dengan menggunakan wordcloud, kata yang memiliki frekuensi kemunculan lebih tinggi akan ditampilkan dengan ukuran yang lebih besar, sehingga memudahkan dalam mengidentifikasi kata dominan.



Gambar 4.28 *WordCloud* Sentimen Positif

Berdasarkan hasil visualisasi pada Gambar 4.13, kata-kata yang paling dominan pada sentimen positif antara lain “yang”, “dan”, “untuk”, “gizi”, “anak”, “makan”, dan “siang”. Kemunculan kata-kata seperti “gizi” dan “anak” menunjukkan bahwa sebagian besar opini positif berkaitan dengan manfaat program terhadap pemenuhan kebutuhan gizi, khususnya bagi anak-anak. Selain itu, kata “makan siang” dan “gratis” (yang juga terlihat meskipun tidak paling dominan) mengindikasikan adanya persepsi positif terhadap program bantuan

makanan. Hal ini menunjukkan bahwa pengguna cenderung memberikan respon yang baik terhadap program yang dinilai bermanfaat bagi masyarakat.

Gambar 4.29 *WordCloud* Sentimen Negatif

Sementara itu, pada Gambar 4.14 wordcloud sentimen negatif, kata-kata yang dominan antara lain “yang”, “dan”, “tidak”, “ini”, “saya”, “untuk”, dan “anak”. Kemunculan kata “tidak” dengan ukuran yang cukup besar menjadi indikator utama adanya opini negatif, yang biasanya berkaitan dengan penolakan, ketidakpuasan, atau kritik terhadap program. Selain itu, kata seperti “ini” dan “saya” menunjukkan bahwa banyak komentar bersifat subjektif atau berupa opini pribadi pengguna.

Jika dibandingkan antara kedua wordcloud, terlihat bahwa terdapat beberapa kata yang muncul pada kedua sentimen, seperti “yang”, “dan”, dan “anak”. Hal ini menunjukkan bahwa topik pembahasan antara sentimen positif dan negatif masih berada pada konteks yang sama, yaitu terkait program yang dianalisis. Namun, perbedaan utama terletak pada kata-kata yang mengandung makna evaluatif, di mana sentimen positif lebih banyak mengandung kata yang bersifat

mendukung atau menggambarkan manfaat, sedangkan sentimen negatif lebih banyak mengandung kata yang menunjukkan penolakan atau kritik.

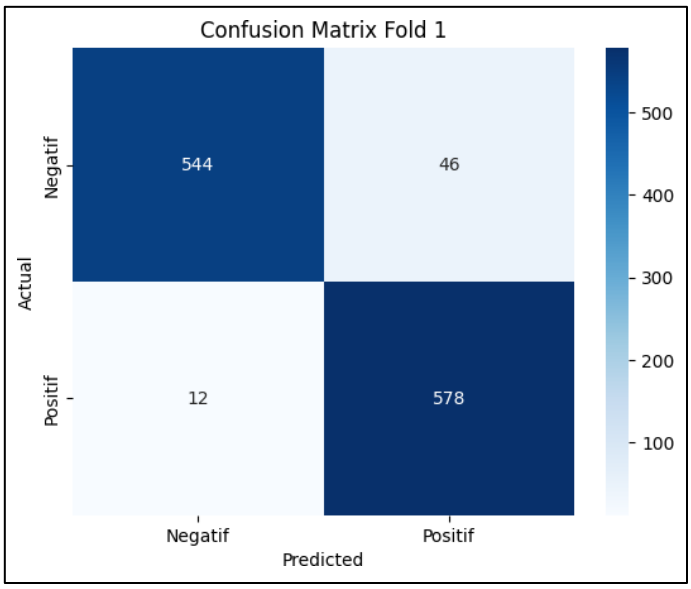
4.5 K-Fold Cross Validation

Pada tahap ini, dilakukan evaluasi model menggunakan metode K-Fold Cross Validation untuk mengetahui tingkat kestabilan dan kemampuan generalisasi model terhadap data yang berbeda. Metode ini digunakan untuk memastikan bahwa model tidak hanya bekerja baik pada satu pembagian data saja, tetapi juga konsisten ketika diuji pada beberapa variasi data.

Dalam penelitian ini digunakan 5-Fold Cross Validation, dimana dataset dibagi menjadi 5 bagian (fold). Setiap fold secara bergantian digunakan sebagai data uji, sementara 4 fold lainnya digunakan sebagai data latih. Proses ini dilakukan sebanyak 5 kali hingga seluruh data pernah menjadi data uji.

Step	Training Loss
50	0.554292
100	0.337049
150	0.336127
200	0.252069
250	0.244924
300	0.222363
350	0.103446
400	0.100333
450	0.077449
500	0.062183
550	0.101635

Gambar 4.30 Hasil Fold 1



Gambar 4.31 Confusion Matrix Fold 1

Berdasarkan hasil pengujian menggunakan metode K-Fold Cross Validation pada fold pertama, model berhasil mengklasifikasikan 544 data negatif dengan benar sebagai negatif (*true negative*) dan 578 data positif dengan benar sebagai positif (*true positive*). Hal ini menunjukkan bahwa model sudah mampu mengenali kedua kelas dengan cukup akurat.

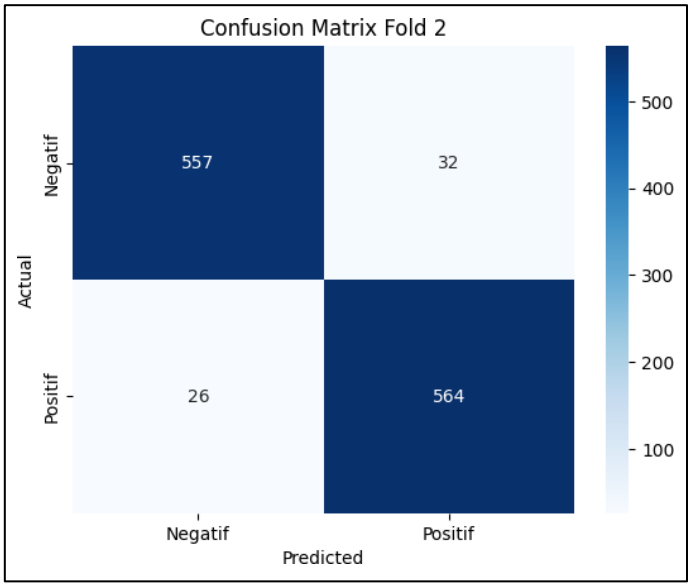
Adapun kesalahan klasifikasi yang terjadi terdiri dari 46 data negatif yang diprediksi sebagai positif (*false positive*) dan 12 data positif yang diprediksi sebagai negatif (*false negative*). Jumlah false negative yang relatif kecil menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mendeteksi data positif, yang juga sejalan dengan nilai recall yang tinggi.

Model memperoleh accuracy sebesar 0.9508, yang menunjukkan bahwa sekitar 95% data berhasil diklasifikasikan dengan benar. Nilai precision sebesar 0.9263 mengindikasikan bahwa sebagian besar prediksi positif yang dihasilkan

model sudah tepat. Sementara itu, recall sebesar 0.9797 dan nilai F1-score sebesar 0.9522.

Step	Training Loss
50	0.460277
100	0.358251
150	0.277481
200	0.286684
250	0.230732
300	0.185464
350	0.085258
400	0.083136
450	0.040361
500	0.070675
550	0.065600

Gambar 4.32 Hasil Fold 2



Gambar 4.33 Confusion Matrix Fold 2

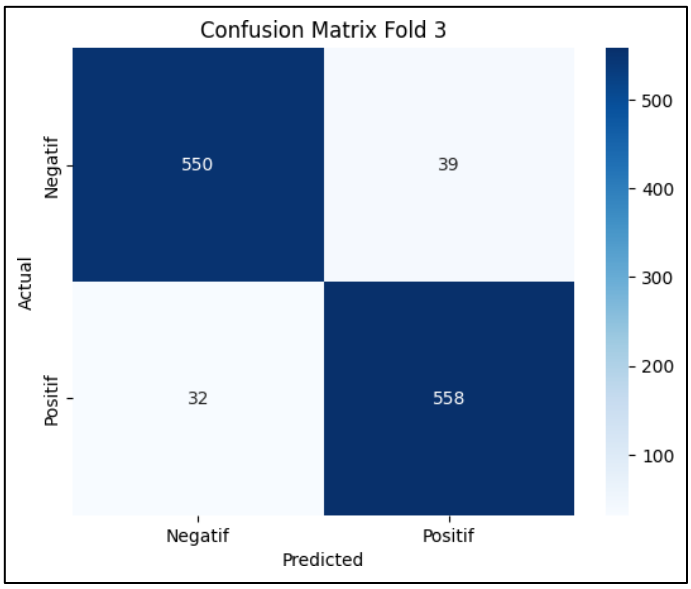
Berdasarkan hasil pengujian pada fold kedua, model berhasil mengklasifikasikan 557 data negatif dengan benar sebagai negatif (*true negative*) dan 564 data positif dengan benar sebagai positif (*true positive*). Hal ini menunjukkan bahwa kemampuan model dalam mengenali kedua kelas masih terjaga dengan baik.

Adapun kesalahan klasifikasi yang terjadi terdiri dari 32 data negatif yang diprediksi sebagai positif (*false positive*) dan 26 data positif yang diprediksi sebagai negatif (*false negative*). Jika dibandingkan dengan fold pertama, jumlah false positive pada fold kedua mengalami penurunan, yang menunjukkan adanya peningkatan dalam ketepatan model saat mengklasifikasikan data negatif.

Model memperoleh accuracy sebesar 0.9508, yang menunjukkan tingkat akurasi yang sama dengan fold sebelumnya. Nilai precision sebesar 0.9463 mengalami peningkatan dibandingkan fold pertama. Sementara itu, recall sebesar 0.9559 dan nilai F1-score sebesar 0.9511.

Step	Training Loss
50	0.445159
100	0.374035
150	0.324852
200	0.274038
250	0.253261
300	0.207548
350	0.076045
400	0.073702
450	0.056555
500	0.071732
550	0.076079

Gambar 4.34 Hasil Fold 3



Gambar 4.35 Confusion Matrix Fold 3

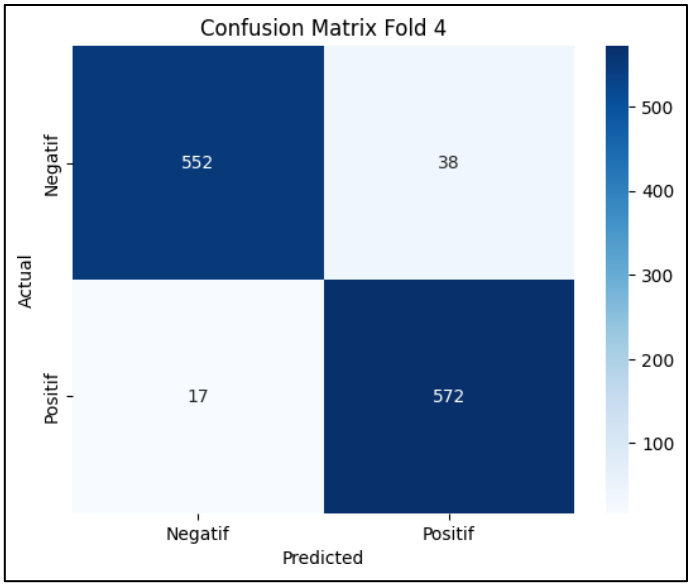
Berdasarkan hasil pengujian pada fold ketiga, model berhasil mengklasifikasikan 550 data negatif dengan benar sebagai negatif (*true negative*) dan 558 data positif dengan benar sebagai positif (*true positive*). Hal ini menunjukkan bahwa model tetap mampu mengenali kedua kelas dengan cukup baik.

Adapun kesalahan klasifikasi yang terjadi terdiri dari 39 data negatif yang diprediksi sebagai positif (*false positive*) dan 32 data positif yang diprediksi sebagai negatif (*false negative*). Jumlah kesalahan pada fold ini sedikit lebih tinggi dibandingkan fold pertama dan kedua, yang menunjukkan adanya penurunan performa model pada pembagian data ini.

Model memperoleh accuracy sebesar 0.9398, yang merupakan nilai terendah dibandingkan fold sebelumnya. Nilai precision sebesar 0.9347 masih tergolong baik. Sementara itu, recall sebesar 0.9458 dan nilai F1-score sebesar 0.9402.

Step	Training Loss
50	0.442228
100	0.408414
150	0.297079
200	0.229143
250	0.268751
300	0.191054
350	0.098582
400	0.099920
450	0.075388
500	0.073694
550	0.100334

Gambar 4.36 Hasil Fold 4



Gambar 4.37 Confusion Matrix Fold 4

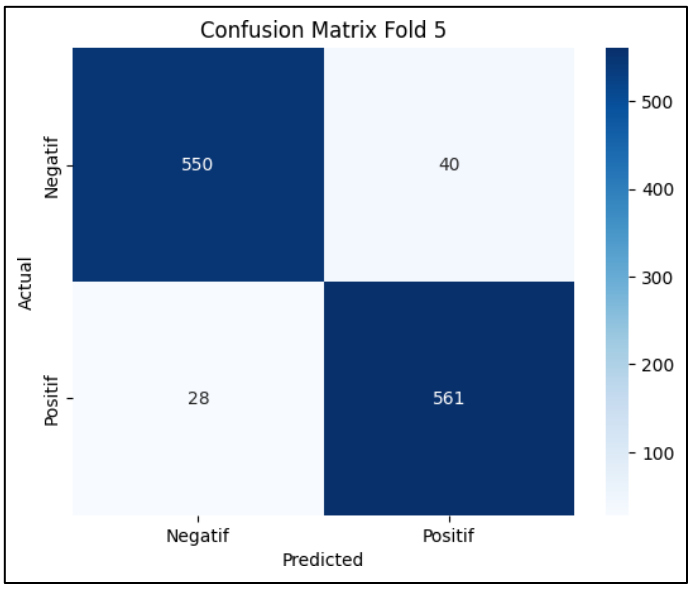
Berdasarkan hasil pengujian pada fold keempat, model berhasil mengklasifikasikan 552 data negatif dengan benar sebagai negatif (*true negative*) dan 572 data positif dengan benar sebagai positif (*true positive*).

Adapun kesalahan klasifikasi yang terjadi terdiri dari 38 data negatif yang diprediksi sebagai positif (*false positive*) dan 17 data positif yang diprediksi sebagai negatif (*false negative*). Jumlah false negative yang relatif rendah menunjukkan bahwa model tetap sangat baik dalam mendeteksi data positif.

Model memperoleh accuracy sebesar 0.9534, yang merupakan salah satu nilai tertinggi dibandingkan fold sebelumnya. Nilai precision sebesar 0.9377, recall sebesar 0.9711 dan nilai F1-score sebesar 0.9541.

Step	Training Loss
50	0.467792
100	0.353355
150	0.310578
200	0.285052
250	0.255109
300	0.205672
350	0.062644
400	0.090929
450	0.077332
500	0.067428
550	0.087113

Gambar 4.38 Hasil Fold 5



Gambar 4.39 Confusion Matrix Fold 5

Berdasarkan hasil pengujian pada fold kelima, model berhasil mengklasifikasikan 550 data negatif dengan benar sebagai negatif (true negative) dan 561 data positif dengan benar sebagai positif (*true positive*).

Adapun kesalahan klasifikasi terdiri dari 40 data negatif yang diprediksi sebagai positif (*false positive*) dan 28 data positif yang diprediksi sebagai negatif (*false negative*). Jumlah kesalahan ini sedikit lebih tinggi dibandingkan fold keempat, yang menyebabkan penurunan performa pada fold ini.

Model memperoleh accuracy sebesar 0.9423, yang masih tergolong tinggi, meskipun lebih rendah dibandingkan fold sebelumnya. Nilai precision sebesar 0.9334, sementara recall sebesar 0.9525 dan nilai F1-score sebesar 0.9429.

4.6 Tampilan Antarmuka

Setelah proses pelatihan dan evaluasi selesai dilakukan, model IndoBERT dengan performa terbaik pada Skenario 10 diimplementasikan ke dalam sebuah aplikasi berbasis web. Implementasi ini bertujuan untuk menguji penerapan model

secara langsung serta memudahkan pengguna dalam melakukan analisis sentimen terhadap komentar yang berkaitan dengan Program Makan Bergizi Gratis (MBG).

Gambar 4.40 Tampilan Antarmuka 1

Gambar 4.41 Tampilan Antarmuka 2

Antarmuka sistem dirancang sederhana agar mudah digunakan. Pengguna cukup memasukkan teks komentar pada kolom yang tersedia, kemudian menekan tombol Analisis Sentimen. Selanjutnya, sistem akan memproses teks masukan menggunakan model IndoBERT yang telah dilatih untuk menentukan kategori sentimen komentar tersebut.

Hasil analisis ditampilkan dalam bentuk label sentimen, yaitu positif atau negatif, disertai nilai keyakinan (confidence score) yang menunjukkan tingkat kepercayaan model terhadap hasil prediksi. Selain itu, sistem juga menampilkan probabilitas masing-masing kelas sentimen dalam bentuk visualisasi batang sehingga pengguna dapat melihat perbandingan tingkat kemungkinan antara sentimen positif dan negatif. Implementasi ini menunjukkan bahwa model yang telah dikembangkan dapat digunakan secara praktis untuk melakukan klasifikasi sentimen secara otomatis terhadap opini masyarakat terkait Program Makan Bergizi Gratis.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat diketahui bahwa sentimen masyarakat terhadap program Makan Bergizi Gratis (MBG) di media sosial cenderung didominasi oleh sentimen negatif. Hal tersebut ditunjukkan dari hasil proses *labelling* data yang memperoleh 2948 data sentimen negatif dan 1256 data sentimen positif. Dominasi sentimen negatif tersebut umumnya berkaitan dengan kritik masyarakat terhadap pelaksanaan program, kualitas makanan, serta proses distribusi di lapangan. Meskipun demikian, masih ditemukan sentimen positif yang menunjukkan adanya dukungan masyarakat terhadap tujuan program dalam meningkatkan pemenuhan gizi masyarakat. Selain itu, hasil penelitian menunjukkan bahwa algoritma IndoBERT mampu memberikan performa yang sangat baik dalam proses klasifikasi sentimen teks berbahasa Indonesia. Hal ini dibuktikan melalui pengujian pada 12 skenario dengan kombinasi rasio data, *batch size*, dan *learning rate* yang berbeda.

Dari seluruh skenario yang diuji, performa terbaik diperoleh pada skenario 10 dengan rasio data 80:20, *batch size* 16, dan *learning rate* 0.00003. Pada skenario tersebut, model berhasil memperoleh *accuracy* sebesar 0.9585, *precision* sebesar 0.9442, *recall* sebesar 0.9746, dan *F1-score* sebesar 0.9591. Hasil tersebut menunjukkan bahwa IndoBERT memiliki kemampuan yang baik dalam memahami konteks bahasa Indonesia serta mampu melakukan klasifikasi sentimen secara akurat dan seimbang. Hasil penelitian juga menunjukkan bahwa parameter yang

digunakan memberikan pengaruh terhadap performa model. Penggunaan rasio data 80:20 cenderung menghasilkan performa yang lebih baik karena model memperoleh data pelatihan yang lebih banyak. *Batch size* 16 juga memberikan hasil yang lebih optimal dibandingkan *batch size* 32 karena proses pembelajaran model menjadi lebih detail, sedangkan *learning rate* sebesar 0.00003 mampu membantu proses *fine-tuning* model menjadi lebih optimal dibandingkan *learning rate* 0.00002 pada sebagian besar skenario pengujian.

5.2 Saran

Penelitian selanjutnya dapat menggunakan jumlah data yang lebih besar dan sumber data yang lebih beragam agar hasil analisis sentimen dapat merepresentasikan opini masyarakat secara lebih luas dan akurat. Selain itu, penelitian selanjutnya juga dapat menambahkan kategori sentimen netral sehingga klasifikasi sentimen tidak hanya terbatas pada sentimen positif dan negatif saja, tetapi dapat menggambarkan kondisi opini masyarakat secara lebih lengkap.

Penelitian berikutnya dapat melakukan perbandingan dengan model lain, seperti LSTM, SVM, IndoBERTweet, maupun model transformer lainnya untuk mengetahui metode yang memiliki performa paling optimal dalam klasifikasi sentimen bahasa Indonesia. Selain itu, penggunaan metode balancing data lain selain *Random Oversampling* juga dapat dipertimbangkan agar dapat mengurangi kemungkinan terjadinya duplikasi data selama proses pelatihan model.

DAFTAR PUSTAKA

Al-kadzim, M. G. (2024). *Analisis Perubahan Sentimen Publik di Media Sosial X terhadap Konflik Palestina-Israel Menggunakan Model IndoBERT*. *4*(2), 1167–1174.

Andhika, F. R., Witanti, W., & Sabrina, P. N. (2025). *Analisis Sentimen Menggunakan Metode IndoBERT pada Ulasan Aplikasi Zoom Menggunakan Fitur Ekstrasi GloVe*. https://doi.org/10.47002/metik.v9i2.1098

Apriani, R., & Gustian, D. (2019). *ANALISIS SENTIMEN DENGAN NAÏVE BAYES TERHADAP KOMENTAR APLIKASI TOKOPEDIA*. *6*(1).

Fadilah, G. T., Muflikhah, L., & Perdana, R. S. (2025). *ANALISIS SENTIMEN PRODUK HIJAB PADA E-COMMERCE TOKOPEDIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR*. *9*(2), 1–9.

Jeevallucas, P., Surya, J., Wibowo, A., Siang, J. J., Informasi, S., Informasi, F. T., Kristen, U., & Wacana, D. (2025). *ANALISIS SENTIMEN PENGGUNA TERHADAP AKUN X / TWITTER RESMI " DANA " DENGAN ALGORITMA INDOBERT*. *8*(November 2024), 549–559.

Karim, H. F., & Wibowo, A. P. (2025). *Kinerja Metode Fine-Tuning IndoBERT untuk Klasifikasi Emosi Multi-Kelas pada Teks Informal Bahasa Indonesia*. *6*(1), 63–74. https://doi.org/10.47065/bulletincsr.v6i1.850

Karim, R. A. (2024). *Sentiment Analysis of YouTube Users on Blackpink Kpop Group Using IndoBERT*. *8*(2), 233–245.

Kusoema, W. J., Ibrahim, I., Studi, P., Informatika, T., Bandung, K., & Barat, J. (2025). *Analisis Sentimen dalam Kasus Korupsi PT . Pertamina menggunakan Metode IndoBERT dan RCNN Sentiment Analysis on the PT Pertamina Corruption Case using IndoBERT and RCNN Methods*. *14*, 2246–2257.

Manoppo, M. R., Kolang, I. C., Fiat, D. N., Michelly, R., & Mawara, C. (2025). *ANALISIS SENTIMEN PUBLIK DI MEDIA SOSIAL TERHADAP KENAIKAN PPN 12 % DI INDONESIA MENGGUNAKAN INDOBERT ANALYSIS OF PUBLIC SENTIMENT ON SOCIAL MEDIA TOWARDS THE 12 % PPN INCREASE IN INDONESIA USING INDOBERT*. *4*(2), 152–163.

Merdiansah, R., & Ridha, A. A. (2024). *Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT*. *7*, 221–228.

Pratama, A. Y., Sanjaya, G. A., Lubis, N. K., & Aditya, M. R. (2025). *Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial*. https://doi.org/10.47002/metik.v9i1.1055

Sarimole, F. M. (2024). *Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine*. *5*(3), 783–790.

Triningsih, E., Afdal, M., Permana, I., & Rozanda, N. E. (2025). *Analisis Sentimen Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning Pada Sosial Media X*. *6*(4), 2240–2250. https://doi.org/10.47065/bits.v6i4.6534

Verena, K., Toy, S., Sari, Y. A., & Cholissodin, I. (2021). *Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus : Opini Masyarakat mengenai Kebijakan New Normal*

Waskito, A., & Danang, D. (2019). *Analisis Sentimen Terhadap Pemilihan Presiden Indonesia 2019 Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Lukito Agung Waskito , Kemas Muslim Lhaksmana , Danang Triantoro Murdiansyah*. *6*(2), 9753–9765.

Widowati, T. T., Komputer, F. I., Studi, P., Informatika, T., Buana, U. M., Sadikin, M., Komputer, F. I., Studi, P., Informatika, T., & Buana, U. M. (2020). *ANALISIS SENTIMEN TWITTER TERHADAP TOKOH PUBLIK DENGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR*. *11*(2).