ANALISIS SENTIMEN PROGRAM PEMERINTAH MAKAN BERGIZI GRATIS MENGGUNAKAN METODE *INDOBERT TRANSFORMER*

SKRIPSI

DISUSUN OLEH:

HANDHALLA AKASYAH AL-KAUTSAR PANJAITAN

NPM. 2109020012

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN 2025

ANALISIS SENTIMEN PROGRAM PEMERINTAH MAKAN BERGIZI GRATIS MENGGUNAKAN METODE *INDOBERT TRANSFORMER*

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) dalam Program Studi Teknologi Informasi pada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara

HANDHALLA AKASYAH AL-KAUTSAR PANJAITAN
NPM.2109020012

PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN 2025

LEMBAR PENGESAHAN

Judul Skripsi	:	Analisis Sentimen Program Pemerintah Makan Bergizi Gratis Menggunakan Metode *Indobert Transformer*
Nama Mahasiswa	:	Handhalla Akasyah Alkautsar Panjaitan
NPM	:	2109020012
Program Studi	:	Teknologi Informasi

Menyetujui
Dosen Pembimbing

(Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom.)
NIDN. 0117088902

Ketua Program Studi	Dekan
	
(Fatma Sari Hutagalung S.Kom., M.Kom.)	**(Dr. Al-Khowarizmi, S.Kom., M.Kom.)**
NIDN. 0117019301	**NIDN. 0127099201**

PERNYATAAN ORISINALITAS

ANALISIS SENTIMEN PROGRAM PEMERINTAH MAKAN BERGIZI GRATIS MENGGUNAKAN METODE *INDOBERT TRANSFORMER*

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 1 Agustus 2025

Yang membuat pernyataan

Handhalla Akasyah Alkautsar Panjaitan

NPM. 2109020012

PERNYATAAN PERSETUJUAN PUBLIKASI

KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: Handhalla Akasyah Alkautsar Panjaitan
NPM	: 2109020012
Program Studi	: Teknologi Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif *(Non-Exclusive Royalty free Right)* atas penelitian skripsi saya yang berjudul:

ANALISIS SENTIMEN PROGRAM PEMERINTAH MAKAN BERGIZI GRATIS MENGGUNAKAN METODE *INDOBERT TRANSFORMER*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 8 September 2025

Yang membuat pernyataan

Handhalla Akasyah Alkautsar Panjaitan

NPM. 2109020012

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap	: Handhalla Akasyah Alkautsar Panjaita
Tempat dan Tanggal Lahir	: Medan, 17 Oktober 2003
Alamat Rumah	: Jl Sederhana gang raya 2 pasar 7 Tembung
Telepon/Faks/HP	: 081274465574
E-mail	: handalaakasah7@gmail.com
Instansi Tempat Kerja	: -
Alamat Kantor	: -

DATA PENDIDIKAN

SD	: SD Nurul Hasanah	TAMAT: 2015
SMP	: SMP Negeri 6 Medan	TAMAT: 2018
SMA	: SMA Negeri 5 Medan	TAMAT: 2021

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh

Alhamdulillah, puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, serta nikmat kesehatan dan kesempatan yang telah diberikan, sehingga penulis dapat menyelesaikan penyusunan skripsi ini yang berjudul "Implementasi *Machine Learning* Dengan Model LSTM Untuk Forecasting Harga Saham Menggunakan Data" sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknologi Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara.

Penulis menyadari bahwa keberhasilan dalam penyusunan skripsi ini tidak terlepas dari dukungan dan bantuan berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

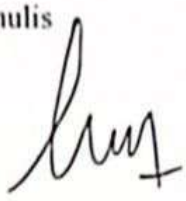
1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU
3. Ibu Fatma Sari Hutagaulung, S.Kom., M.Kom., selaku Ketua Program Studi Teknologi Informasi.
4. Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom.,Selaku Dosen Pembimbing Akademik Penulis.
5. Bapak Muhammad Basri, S.Kom., M.Kom., selaku Sekretaris Program Studi Teknologi Informasi.
6. Seluruh dosen dan staf akademik di lingkungan Fakultas Ilmu Komputer dan Teknologi Informasi, khususnya Program Studi Teknologi Informasi, yang telah memberikan ilmu dan pengalaman berharga selama masa studi.
7. Kedua orang tua tercinta, Bapak Burhanuddin Panjaitan dan Ibu Muliati, yang selalu memberikan doa, semangat, cinta, serta dukungan moral dan material tanpa henti.

8. Keluarga Besar Saya yang selalu memberikan doa, nasihat, serta dukungan kepada penulis selama proses pendidikan hingga penyusunan skripsi ini.
9. Rekan-rekan seperjuangan di grup ojol dan kkn aek natolu 2024 yang telah menjadi teman diskusi sekaligus motivasi selama menempuh studi di UMSU.
10. Terimakasih banyak kepada orang terdekat saya Nur Azizah Sirait, S.Pd yang senantiasa memberikan dukungan moral, motivasi, serta semangat di setiap langkah penulis dalam menyelesaikan tugas akhir ini. Kehadiranmu menjadi penguat di saat penulis hampir menyerah dan penyemangat di kala penulis lelah.
11. Teman-teman kelas TI A1 Stambuk 21 yang sudah menemani penulis sedari awal kegiatan perkuliahan hingga akhir perkuliahan terimakasih sebanyak-banyaknya.
12. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu dan memberikan dukungan secara langsung maupun tidak langsung dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun untuk perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi referensi yang berguna, khususnya dalam pengembangan sistem prediksi harga saham berbasis deep learning.

Medan, 08 September 2025

Penulis



Handhalla Akasyah Alkautsar Panjaitan

ANALISIS SENTIMEN PROGRAM PEMERINTAH MAKAN BERGIZI GRATIS MENGGUNAKAN METODE *INDOBERT TRANSFORMER*

ABSTRAK

Masalah gizi buruk dan stunting masih menjadi tantangan serius di Indonesia. Sebagai upaya strategis, pemerintah meluncurkan Program Makan Bergizi Gratis (MBG) pada tahun 2025 yang ditujukan untuk anak-anak sekolah dan ibu hamil. Program ini memunculkan berbagai respon masyarakat di media sosial, baik positif maupun negatif. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program MBG dengan menggunakan metode IndoBERT Transformer, sebuah model deep learning yang dioptimalkan untuk bahasa Indonesia. Data penelitian diambil dari dataset Kaggle yang berisi 656 ulasan publik terkait program MBG. Proses penelitian meliputi preprocessing data (case folding, cleaning, normalisasi), tokenisasi, embedding, pelatihan, serta evaluasi model. Hasil pengujian menunjukkan bahwa model IndoBERT mampu mengklasifikasikan sentimen dengan akurasi sebesar 85%, precision 0.92 pada kelas negatif, serta recall 0.68 pada kelas positif. Temuan ini menunjukkan bahwa IndoBERT efektif dalam menangkap konteks bahasa Indonesia pada opini publik, meskipun terdapat tantangan pada data yang tidak seimbang. Penelitian ini diharapkan dapat memberikan masukan bagi pemerintah dalam mengevaluasi dan mengembangkan program MBG, sekaligus memperkaya literatur mengenai penerapan Transformer untuk analisis sentimen kebijakan publik di Indonesia**.**

Kata Kunci: Analisis Sentimen, IndoBERT, Transformer, Makan Bergizi Gratis, Media Sosial

SENTIMENT ANALYSIS OF THE GOVERNMENST'S FREE NUTRITIOUS MEAL PROGRAM USING THE INDOBERT TRANSFORMER METHOD

ABSTRACT

Malnutrition and stunting remain critical challenges in Indonesia. As a strategic effort, the government launched the Free Nutritious Meal (MBG) Program in 2025, targeting school children and pregnant women. This program has generated diverse public responses on social media, both positive and negative. This study aims to analyze public sentiment toward the MBG program using the IndoBERT Transformer, a deep learning model optimized for the Indonesian language. The dataset, obtained from Kaggle, consists of 656 public reviews related to the MBG program. The research process includes data preprocessing (case folding, cleaning, normalization), tokenization, embedding, model training, and evaluation. Experimental results show that IndoBERT achieved 85% accuracy, with 0.92 precision on the negative class and 0.68 recall on the positive class. These findings indicate that IndoBERT is effective in capturing the nuances of the Indonesian language in public opinion, although challenges remain due to imbalanced data. This study is expected to provide valuable insights for the government in evaluating and improving the MBG program, while also contributing to the literature on the application of Transformers for sentiment analysis in public policy contexts in Indonesia.

Keywords: Sentiment Analysis, IndoBERT, Transformer, Free Nutritious Meal Program, Social Media

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB I
PENDAHULUAN

1.1. Latar Belakang Masalah

Masalah gizi buruk dan stunting masih menjadi tantangan serius di Indonesia. Berdasarkan penelitian Survei Status Gizi Indonesia (SSGI) tahun 2022, prevalensi stunting pada balita mencapai 21,6%, menunjukkan bahwa satu dari lima anak mengalami gangguan pertumbuhan akibat kekurangan gizi kronis (Mustaqim et al., 2024). Gizi buruk tidak hanya meningkatkan angka morbiditas dan mortalitas, tetapi juga menurunkan produktivitas serta menghambat pertumbuhan sel-sel otak, yang berdampak pada kualitas sumber daya manusia di masa depan (Hendraswari, 2023).

Sebagai langkah strategis untuk mengatasi permasalahan ini, pemerintah Indonesia meluncurkan program Makan Bergizi Gratis (MBG) pada 6 Januari 2025. Program ini bertujuan untuk meningkatkan asupan gizi bagi anak-anak sekolah dan ibu hamil, dengan target menjangkau sekitar 83 juta penerima manfaat dan anggaran awal sebesar Rp71 triliun . Selain itu, program ini juga diharapkan dapat mendukung pertumbuhan ekonomi lokal melalui pelibatan Usaha Mikro, Kecil, dan Menengah (UMKM) dalam penyediaan bahan pangan local (Report, 2024)

Pelaksanaan program MBG mendapat perhatian luas dari masyarakat, terutama melalui media sosial seperti platform X (sebelumnya dikenal sebagai Twitter). Platform ini menjadi saluran utama bagi publik untuk menyampaikan pandangan mereka tentang kebijakan dan program pemerintah. Data dari platform X yang tersedia di Kaggle dapat memberikan wawasan tentang bagaimana masyarakat memandang program ini. Analisis sentimen terhadap data ini penting

untuk memahami persepsi publik dan memberikan masukan konstruktif bagi perbaikan dan pengembangan program MBG.

Analisis sentimen merupakan metode yang efektif untuk mengkaji opini publik melalui data teks dari media sosial. Namun, tantangan muncul karena karakteristik bahasa Indonesia yang kompleks, termasuk penggunaan bahasa gaul, singkatan, dan struktur kalimat yang beragam. Model tradisional seperti Support Vector Machine (SVM) sering kali kurang mampu menangkap konteks dan nuansa dalam bahasa Indonesia secara efektif. Sebagai solusi, pendekatan berbasis transformer seperti IndoBERT telah dikembangkan untuk memahami konteks bahasa Indonesia dengan lebih baik.

IndoBERT adalah model pra-latih berbasis arsitektur BERT yang dioptimalkan khusus untuk bahasa Indonesia. Model ini telah menunjukkan performa unggul dalam berbagai tugas pemrosesan bahasa alami. Misalnya, penelitian oleh (Jayadianti et al., 2022) menunjukkan bahwa kombinasi IndoBERT dengan Recurrent Convolutional Neural Network (RCNN) mencapai akurasi sebesar 95,16% dalam analisis sentimen ulasan berbahasa Indonesia . Hasil ini menunjukkan bahwa IndoBERT mampu menangkap konteks dan nuansa bahasa Indonesia dengan baik.

Selain itu, penelitian oleh (Hidayat & Pramudita, 2024) menggunakan IndoBERT untuk menganalisis sentimen terhadap pembelajaran daring pasca pandemi COVID-19. Model ini berhasil mencapai akurasi sebesar 87%, dengan precision sebesar 87%, recall sebesar 91%, dan F1-score sebesar 89% . Studi ini menunjukkan bahwa IndoBERT efektif dalam mengklasifikasikan sentimen pada data teks berbahasa Indonesia dalam konteks Pendidikan.

Melihat keberhasilan IndoBERT dalam berbagai penelitian dapat diharapkan memperoleh pemahaman yang lebih mendalam mengenai persepsi publik terhadap program tersebut, sehingga dapat memberikan masukan yang konstruktif bagi perumusan dan pelaksanaan kebijakan pemerintah di masa mendatang.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana penerapan metode IndoBERT Transformer dalam analisis sentimen terhadap program Makan Bergizi Gratis berdasarkan data dari media sosial?
2. Bagaimana cara kerja metode IndoBERT Transformer dalam mengklasifikasikan sentimen masyarakat terhadap program Makan Bergizi Gratis?
3. Seberapa tinggi tingkat akurasi metode IndoBERT Transformer dalam mengklasifikasikan sentimen masyarakat terhadap program Makan Bergizi Gratis berdasarkan data dari platform Kaggle?

1.3. Batasan Masalah

Agar penelitian ini lebih terfokus dan mendapatkan hasil yang lebih akurat, terdapat beberapa Batasan masalah yang ditetapkan:

1. Data Data yang digunakan dalam penelitian ini berasal dari dataset terbuka di Kaggle. Dataset tersebut memuat opini masyarakat mengenai program Makan Bergizi Gratis yang diambil dari platform media sosial.
2. Penelitian hanya mengklasifikasikan *dua kelas sentimen* (positif & negatif), tidak mencakup netral.

3. Metode yang digunakan untuk klasifikasi adalah metode IndoBERT Transformer.
4. Data yang dianalisis berupa teks opini atau ulasan dalam bahasa Indonesia.

1.4. Tujuan Penelitian

1. Mengidentifikasi dan menganalisis sentimen masyarakat terhadap program Makan Bergizi Gratis melalui data dari Kaggle yang berasal dari media sosial.
2. Mengevaluasi kinerja metode IndoBERT Transformer dalam mengklasifikasi sentimen tersebut
3. Memberikan rekomendasi bagi pemerintah berdasarkan hasil analisis sentimen untuk perbaikan dan pengembangan program Makan Bergizi Gratis.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan gambaran kepada pemerintah mengenai persepsi masyarakat terhadap program Makan Bergizi Gratis, sehingga dapat digunakan sebagai bahan evaluasi dan perbaikan kebijakan
2. Menambah literatur akademik mengenai penerapan metode IndoBERT Transformer dalam analisis sentimen terhadap kebijakan publik di Indonesia.
3. Menyediakan referensi bagi peneliti selanjutnya yang ingin mengkaji analisis sentimen terhadap program pemerintah menggunakan metode Transformer, khususnya IndoBERT

BAB II
LANDASAN TEORI

2.1 Program Makan Bergizi Gratis (MBG)

Program Makan Bergizi Gratis (MBG) merupakan salah satu program prioritas nasional yang diluncurkan oleh pemerintah Indonesia pada 6 Januari 2025. Program ini bertujuan untuk mengatasi masalah malnutrisi dan stunting yang masih menjadi tantangan serius di Indonesia. Menurut data Survei Kesehatan Indonesia 2023, prevalensi stunting di Indonesia mencapai 21,5%, menunjukkan perlunya intervensi yang efektif dalam pemenuhan gizi anak-anak dan ibu hamil. Melalui MBG, pemerintah berupaya memberikan asupan gizi yang memadai kepada kelompok rentan tersebut untuk meningkatkan kualitas sumber daya manusia Indonesia

Pelaksanaan program MBG diawali dengan pemberian makanan bergizi kepada sekitar 19,5 juta anak-anak dan ibu hamil di 26 provinsi di Indonesia. Program ini dirancang untuk menjangkau hingga 83 juta penerima manfaat secara bertahap hingga tahun 2029. Anggaran yang dialokasikan untuk tahap awal program ini mencapai Rp71 triliun, dengan estimasi total biaya sebesar Rp450 triliun hingga akhir periode pelaksanaan. Pemerintah juga melibatkan sekitar 2.000 koperasi dalam penyediaan bahan makanan, seperti beras, sayuran, daging ayam, dan susu, guna mendukung perekonomian lokal dan memastikan keberlanjutan program

Program MBG tidak hanya berfokus pada pemenuhan kebutuhan gizi, tetapi juga memiliki dampak positif terhadap perekonomian nasional. Dengan melibatkan petani, peternak, dan nelayan lokal sebagai pemasok utama bahan makanan, program ini diharapkan dapat meningkatkan pendapatan masyarakat di daerah dan

menciptakan lapangan kerja baru. Menurut estimasi, pelaksanaan program MBG dapat menyerap sekitar 820.000 tenaga kerja dan menyumbang peningkatan pertumbuhan ekonomi Indonesia sekitar 0,1% pada tahun 2025

Untuk memastikan kualitas dan keamanan makanan yang disediakan, pemerintah menetapkan standar gizi yang ketat dan melakukan pengawasan secara berkala. Setiap porsi makanan dirancang untuk memenuhi sepertiga kebutuhan gizi harian penerima manfaat, dengan menu yang disesuaikan berdasarkan ketersediaan bahan pangan lokal. Selain itu, pengelolaan limbah dapur juga menjadi perhatian, dengan penerapan sistem pengolahan limbah yang ramah lingkungan dan penggunaan peralatan penyajian yang dapat digunakan ulang.

Meskipun program MBG memiliki tujuan yang mulia, pelaksanaannya menghadapi berbagai tantangan, termasuk dalam hal logistik dan pendanaan. Beberapa pihak mengkhawatirkan keberlanjutan program ini mengingat besarnya anggaran yang diperlukan dan potensi dampaknya terhadap defisit anggaran negara. Namun, pemerintah berkomitmen untuk mengatasi tantangan tersebut melalui koordinasi dengan berbagai pihak, termasuk pemerintah daerah dan sektor swasta, serta dengan memanfaatkan dana dari Anggaran Pendapatan dan Belanja Negara (APBN) dan sumber lainnya

Secara keseluruhan, Program Makan Bergizi Gratis merupakan langkah strategis pemerintah Indonesia dalam upaya meningkatkan kualitas sumber daya manusia dan mewujudkan visi Indonesia Emas 2045. Dengan memastikan akses terhadap makanan bergizi bagi anak-anak dan ibu hamil, program ini diharapkan dapat menurunkan angka stunting, meningkatkan kesehatan masyarakat, dan mendukung pertumbuhan ekonomi yang inklusif dan berkelanjutan.

Namun, pelaksanaan program MBG tidak lepas dari tantangan. Beberapa kritik muncul terkait keberlanjutan anggaran, logistik distribusi makanan, dan potensi penyalahgunaan dana. Studi oleh (Andin et al., 2025) menyoroti bahwa meskipun program ini memiliki tujuan mulia, diperlukan pengawasan yang ketat dan evaluasi berkelanjutan untuk memastikan efektivitasnya

2.2 Analisis Sentimen

Analisis sentimen adalah cabang dari text mining yang bertujuan untuk mengklasifikasikan opini atau perasaan seseorang terhadap suatu objek atau topik tertentu. Dalam konteks ini, opini biasanya dikategorikan ke dalam sentimen positif, negatif, atau netral. Menurut (Rofiqoh et al., 2017), analisis sentimen berguna untuk memahami opini publik secara otomatis dari teks yang bersifat subjektif, seperti komentar media sosial atau ulasan produk.

Dalam praktiknya, analisis sentimen menggunakan beberapa pendekatan utama, seperti metode berbasis kamus (lexicon-based) dan pembelajaran mesin (machine learning). Pendekatan lexicon-based menggunakan daftar kata-kata dengan nilai sentimen tertentu untuk menilai teks, sedangkan pendekatan machine learning memerlukan data latih berlabel untuk membangun model klasifikasi.

Keberhasilan analisis sentimen sangat bergantung pada kualitas data dan proses pra-pemrosesan teks. Pra-pemrosesan meliputi tahapan pembersihan data, penghilangan stop words, stemming, dan tokenisasi untuk mengubah teks mentah menjadi format yang dapat dianalisis oleh algoritma. Menurut (Musfiroh et al., 2021), proses ini juga membantu mengurangi kebisingan data dan meningkatkan akurasi klasifikasi sentimen.

2.3 Machine Learning

Machine Learning (ML) merupakan bagian dari ilmu komputer yang menggunakan teknik statistik, matematika, dan kecerdasan buatan untuk mengekstraksi pola dan pengetahuan dari data berukuran besar. Machine Learning memungkinkan komputer untuk belajar dari data tanpa harus diprogram secara eksplisit, sehingga dapat melakukan prediksi atau klasifikasi berdasarkan pengalaman sebelumnya.

Dalam konteks data mining, Machine Learning berperan sebagai metode cerdas untuk menemukan pola tersembunyi dari basis data yang besar.

Terdapat beberapa pendekatan dalam ML, yaitu:

1. Supervised Learning – model dilatih menggunakan data berlabel untuk melakukan klasifikasi atau regresi.
2. Unsupervised Learning – model mencari pola atau pengelompokan data tanpa label, misalnya clustering (K-Means).
3. Reinforcement Learning – model belajar melalui interaksi dengan lingkungan berdasarkan reward dan punishment.

Dalam penelitian (Maulana & Al-Khowarizmi, 2021) , Machine Learning diterapkan pada analisis data evaluasi pembelajaran mahasiswa dengan metode clustering (K-Means++), yang termasuk dalam unsupervised learning. Tujuannya adalah untuk mengelompokkan data evaluasi berdasarkan kesamaan karakteristik, sehingga menghasilkan wawasan baru terkait kualitas pembelajaran.

2.4 Deep Learning

Deep Learning merupakan salah satu cabang dari machine learning yang memanfaatkan jaringan saraf tiruan (artificial neural networks) dengan arsitektur

yang lebih dalam (memiliki banyak lapisan) untuk mengekstraksi fitur kompleks dari data. Pendekatan ini memungkinkan komputer untuk secara otomatis mempelajari representasi data pada berbagai tingkat abstraksi, sehingga lebih unggul dibandingkan metode pembelajaran tradisional dalam mengenali pola yang kompleks.

Sejak tahun 2010, deep learning mulai banyak digunakan dalam berbagai bidang, termasuk computer vision, pengenalan suara, pemrosesan bahasa alami, dan deteksi objek. Salah satu bentuk populer adalah Convolutional Neural Network (CNN), yang dirancang khusus untuk menganalisis data visual dengan memanfaatkan operasi konvolusi untuk mengekstraksi fitur dari citra.

Menurut (Nasution et al., 2023) arsitektur deep learning seperti CNN, R-CNN, dan Faster R-CNN terdiri atas beberapa lapisan, yaitu convolutional layer, pooling layer, dan fully connected layer. Lapisan-lapisan ini bekerja secara berjenjang untuk mengekstraksi fitur, mereduksi dimensi, hingga menghasilkan prediksi kelas. Proses pelatihan deep learning menggunakan fungsi aktivasi (misalnya ReLU) dan optimisasi loss function untuk meminimalkan kesalahan prediksi.

Keunggulan deep learning adalah kemampuannya belajar secara otomatis dari data mentah dalam jumlah besar, sehingga sangat cocok digunakan dalam penelitian modern, termasuk pada bidang NLP (Natural Language Processing) dengan model transformer seperti BERT dan IndoBERT.

2.5 Natural Language Processing(NLP)

Natural Language Processing (NLP) adalah bidang dalam kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia. Tujuan utama

NLP adalah membuat mesin mampu membaca, memahami, menalar, dan memproses bahasa manusia secara otomatis

Dalam implementasinya, NLP memiliki peran penting pada berbagai aplikasi modern, misalnya chatbot, analisis sentimen, sistem tanya jawab, dan rekomendasi informasi. NLP bekerja dengan serangkaian tahapan, seperti tokenisasi, stemming/lemmatisasi, penghapusan stopwords, analisis morfologi, serta representasi teks dalam bentuk vektor numerik sehingga dapat diproses oleh algoritma machine learning atau deep learning

Penelitian (Mubarok & Tanjung, 2024) menunjukkan bahwa penerapan NLP dalam pengembangan chatbot di lingkungan akademik mampu meningkatkan efisiensi layanan dengan memberikan jawaban otomatis terhadap pertanyaan yang sering diajukan mahasiswa. Proses ini melibatkan pengumpulan data percakapan, preprocessing teks, pelatihan model, dan evaluasi kinerja menggunakan algoritma deep learning, seperti Long Short-Term Memory (LSTM). Hasil pengujian menunjukkan chatbot berbasis NLP yang dikembangkan memiliki tingkat akurasi tinggi (97%), membuktikan kemampuan NLP dalam memahami konteks pertanyaan dan memberikan respons yang relevan.

Dengan demikian, NLP bukan hanya sekadar metode pengolahan bahasa, tetapi juga teknologi kunci yang menjadi fondasi dalam sistem berbasis AI, termasuk analisis sentimen menggunakan model transformer seperti BERT dan IndoBERT.

2.6 Transformer dan Model BERT

Transformer adalah arsitektur model deep learning yang diperkenalkan oleh Vaswani et al. (2017) dan telah merevolusi berbagai tugas dalam pemrosesan bahasa alami (Natural Language Processing/NLP). Keunggulan utama Transformer

terletak pada mekanisme self-attention yang memungkinkan model untuk memahami konteks kata dalam sebuah kalimat secara lebih efektif dibandingkan pendekatan sebelumnya seperti Recurrent Neural Network (RNN) atau Long Short-Term Memory (LSTM). Mekanisme ini memungkinkan Transformer untuk memproses data secara paralel, sehingga meningkatkan efisiensi dan akurasi dalam berbagai aplikasi NLP.

Salah satu implementasi paling terkenal dari arsitektur Transformer adalah BERT (Bidirectional Encoder Representations from Transformers) yang dikembangkan oleh Google. BERT dilatih secara bidirectional, yang berarti model ini mempertimbangkan konteks dari kedua arah (kiri dan kanan) dalam memahami makna sebuah kata dalam kalimat. Pendekatan ini memungkinkan BERT untuk menangkap nuansa makna yang lebih kompleks dalam teks. Di Indonesia, adaptasi BERT untuk bahasa lokal telah dilakukan melalui pengembangan model IndoBERT, yang telah menunjukkan kinerja unggul dalam berbagai tugas NLP berbahasa Indonesia (Amien et al., 2024).

Penelitian terbaru oleh (Rahman et al., 2025) menerapkan model BERT untuk analisis sentimen ulasan pengguna aplikasi Blibli. Dalam penelitian tersebut, BERT digunakan untuk mengklasifikasikan sentimen berdasarkan tiga aspek utama: usable, valuable, dan credible. Hasilnya menunjukkan bahwa konfigurasi terbaik diperoleh pada learning rate 5e-6 dan batch size 128, dengan akurasi mencapai 97,27% dan F1-score sebesar 96,46%. Penelitian ini menegaskan efektivitas BERT dalam memahami dan mengklasifikasikan sentimen dalam teks berbahasa Indonesia.

2.7 IndoBERT Transformer

IndoBERT merupakan model bahasa berbasis arsitektur BERT yang dikembangkan khusus untuk bahasa Indonesia. Model ini dilatih menggunakan korpus besar berbahasa Indonesia, seperti Wikipedia bahasa Indonesia, berita online, dan media sosial, sehingga mampu memahami konteks dan struktur bahasa Indonesia dengan lebih baik dibandingkan model BERT multibahasa. Dengan pelatihan pada data lokal, IndoBERT dapat menghasilkan representasi kontekstual yang lebih akurat untuk berbagai tugas pemrosesan bahasa alami (Natural Language Processing/NLP) dalam bahasa Indonesia.

Penelitian oleh (Nugroho et al., 2025) menerapkan IndoBERT untuk analisis sentimen terhadap ulasan aplikasi Mobile JKN di Google Play Store. Hasil penelitian menunjukkan bahwa IndoBERT mampu mengklasifikasikan sentimen pengguna dengan akurasi tinggi, mengidentifikasi aspek-aspek layanan yang perlu ditingkatkan berdasarkan ulasan pengguna. Penelitian ini menegaskan efektivitas IndoBERT dalam memahami dan mengklasifikasikan sentimen dalam teks berbahasa Indonesia.

Selain itu, penelitian oleh (Alfarobby & Irawan, 2024) menggunakan IndoBERT untuk menganalisis sentimen pengguna terhadap aplikasi transportasi online Gojek dan Grab. Dengan menggabungkan IndoBERT dan pemodelan topik, penelitian ini berhasil mengidentifikasi sentimen pengguna serta topik-topik utama yang dibahas dalam ulasan, memberikan wawasan yang berharga bagi pengembangan layanan.

Lebih lanjut, penelitian oleh (Apriyadi, 2025) membandingkan kinerja IndoBERT dengan Support Vector Machine (SVM) dalam menganalisis sentimen

publik terhadap serangan siber pada Bank Syariah Indonesia. Hasil penelitian menunjukkan bahwa IndoBERT memiliki akurasi dan F1-Score yang lebih tinggi dibandingkan SVM, serta mampu mengidentifikasi topik-topik utama dalam sentimen negatif, seperti masalah pada mobile banking dan penarikan dana.

2.8 Hyperparameter Dalam Pemodelan Transformer

Hyperparameter merupakan parameter yang nilainya ditentukan sebelum proses pelatihan model dimulai, dan sangat berpengaruh terhadap kinerja model berbasis Transformer seperti BERT dan IndoBERT. Berbeda dengan parameter model yang dipelajari selama proses training (misalnya bobot dan bias), hyperparameter harus ditetapkan terlebih dahulu dan menjadi acuan dalam proses pembelajaran.

Penyetelan hyperparameter yang tepat terbukti mampu meningkatkan akurasi model IndoBERT pada tugas analisis sentimen, karena setiap kombinasi hyperparameter dapat memberikan hasil performa yang berbeda pada dataset yang sama (Arifadilah, 2023).

2.9 Goggle Colab

Google Colab (Google Colaboratory) adalah platform komputasi awan berbasis Jupyter Notebook yang dikembangkan oleh Google. Platform ini memungkinkan pengguna untuk menulis dan menjalankan kode Python secara interaktif, terutama dalam pengolahan data, pembelajaran mesin, dan analisis statistik. Google Colab populer di kalangan akademisi, peneliti, dan pengembang karena aksesnya yang mudah, gratis, dan terintegrasi langsung dengan Google Drive. Selain itu, Google Colab menawarkan penggunaan GPU dan TPU secara

cuma-cuma dalam batas waktu tertentu, yang sangat membantu dalam pemrosesan komputasi berat seperti pelatihan model deep learning.

Salah satu keunggulan utama Google Colab adalah kemampuannya untuk menjalankan kode Python dengan GPU tanpa perlu instalasi lokal. Pengguna hanya membutuhkan akun Google dan akses internet untuk memulai proyek mereka. Notebook yang dibuat di Colab dapat disimpan langsung ke Google Drive, sehingga memudahkan akses dan kolaborasi. Selain itu, Google Colab mendukung penggunaan berbagai pustaka populer seperti TensorFlow, PyTorch, Scikit-learn, Pandas, dan Matplotlib, yang sudah terpasang secara default. Hal ini memungkinkan para pengguna untuk langsung fokus pada analisis dan pengembangan tanpa perlu konfigurasi awal.

Selain mendukung analisis data dan pembelajaran mesin, Google Colab juga dilengkapi fitur kolaborasi real-time. Beberapa pengguna dapat mengedit dan menjalankan kode secara bersamaan, mirip dengan Google Docs. Fitur komentar dan riwayat perubahan juga tersedia untuk mendukung kerja tim. Menurut penelitian dari Bakrie University, fitur kolaborasi ini mempermudah diskusi dan revisi pada proyek berbasis Python, terutama dalam penelitian Bersama (Gunawan, 2021).

Namun, Google Colab juga memiliki beberapa keterbatasan yang perlu diperhatikan. Salah satunya adalah batasan waktu penggunaan GPU, di mana sesi akan terputus jika tidak aktif dalam jangka waktu tertentu. Selain itu, ada pula batasan penyimpanan lokal yang hanya sekitar 12 GB. Meskipun demikian, keuntungan yang ditawarkan Google Colab dalam hal akses gratis, dukungan

komputasi awan, dan kemudahan integrasi dengan Google Drive tetap menjadikannya pilihan utama untuk riset dan pengembangan berbasis Python.

2.10 Kaggle

Kaggle adalah platform online yang menyediakan komunitas bagi para data scientist, analis data, dan pengembang machine learning untuk berkolaborasi, belajar, dan berkompetisi dalam berbagai tantangan data. Platform ini dikenal sebagai tempat di mana para profesional dan pemula dapat mengasah kemampuan mereka melalui kompetisi yang diselenggarakan oleh berbagai perusahaan dan organisasi, yang memberikan dataset nyata dan masalah bisnis yang perlu diselesaikan dengan teknik analisis data dan machine learning. Melalui kompetisi ini, peserta bisa mendapatkan hadiah, pengakuan, serta pengalaman berharga yang dapat meningkatkan portofolio mereka.Pengumpulan Data: Mengumpulkan data dari berbagai sumber, seperti sensor listrik atau catatan penggunaan energi.

Selain kompetisi, Kaggle juga menyediakan berbagai sumber daya pembelajaran yang lengkap, seperti tutorial, notebook interaktif (Kaggle Notebooks), dan dataset publik yang bisa digunakan untuk latihan atau proyek pribadi. Fitur notebook memungkinkan pengguna untuk menulis dan menjalankan kode secara langsung di cloud tanpa perlu mengatur lingkungan pemrograman secara lokal. Hal ini mempermudah pengguna dari berbagai latar belakang untuk belajar dan bereksperimen dengan teknik-teknik data science dan machine learning dengan cara yang sangat praktis.

Kaggle juga memiliki komunitas yang sangat aktif, di mana pengguna dapat berdiskusi, berbagi insight, dan saling membantu dalam menyelesaikan masalah terkait data. Forum diskusi ini menjadi salah satu sumber ilmu yang berharga karena

penggunanya datang dari berbagai negara dan latar belakang, sehingga menawarkan perspektif yang beragam. Dengan interaksi yang intens, pengguna dapat memperluas jaringan profesionalnya sekaligus mendapatkan solusi dari berbagai problem yang mereka hadapi dalam proyek atau kompetisi.

Secara keseluruhan, Kaggle merupakan ekosistem yang mendukung pengembangan keterampilan data science dan machine learning secara menyeluruh, mulai dari belajar dasar, mengasah kemampuan lewat kompetisi, hingga membangun portofolio yang dapat membuka peluang karir. Karena itulah, Kaggle menjadi salah satu platform paling populer dan berpengaruh di bidang data science saat ini.

2.11 Python

Python adalah bahasa pemrograman tingkat tinggi yang dirancang untuk kemudahan penggunaan dan pembacaan kode. Dikembangkan pertama kali oleh Guido van Rossum pada akhir tahun 1980-an dan dirilis pada tahun 1991, Python menekankan sintaks yang sederhana dan jelas, sehingga sangat cocok untuk pemula maupun programmer berpengalaman. Bahasa ini bersifat open-source dan memiliki komunitas yang sangat besar, sehingga banyak sekali pustaka (libraries) dan modul yang tersedia untuk berbagai kebutuhan, mulai dari pengolahan data, pengembangan web, hingga kecerdasan buatan.

Salah satu keunggulan utama Python adalah fleksibilitasnya. Python bisa digunakan untuk berbagai jenis aplikasi, termasuk pengembangan aplikasi web dengan framework seperti Django dan Flask, analisis data dengan library seperti Pandas dan NumPy, pembelajaran mesin menggunakan TensorFlow dan Scikit-learn, serta automasi tugas sehari-hari. Python juga dapat dijalankan di berbagai

platform seperti Windows, macOS, dan Linux, sehingga memudahkan pengembangan lintas sistem operasi.

Python menggunakan konsep pemrograman berorientasi objek (OOP), pemrograman prosedural, serta pemrograman fungsional, sehingga programmer bisa memilih gaya pemrograman yang paling sesuai dengan kebutuhan proyeknya. Sintaks Python yang sederhana dan intuitif membuat kode lebih mudah dipahami dan dipelihara. Misalnya, Python tidak menggunakan tanda kurung kurawal untuk blok kode, melainkan mengandalkan indentasi, yang meningkatkan keterbacaan.

Selain itu, Python dikenal karena kemampuannya dalam integrasi dengan bahasa lain seperti C, C++, dan Java. Hal ini memungkinkan pengembang memanfaatkan performa tinggi dari bahasa tersebut sekaligus tetap menikmati kemudahan pengembangan dengan Python. Karena semua kelebihan tersebut, Python semakin populer di berbagai bidang, terutama di dunia pendidikan, data science, pengembangan aplikasi, dan riset teknologi terbaru seperti kecerdasan buatan dan analisis data besar.

2.12 Penelitian Terkait

Penelitian mengenai analisis sentimen menggunakan IndoBERT Transformer telah banyak dilakukan pada berbagai platform media sosial, seperti Twitter, Facebook, dan ulasan di marketplace. Studi-studi tersebut menunjukkan bahwa metode klasifikasi berbasis Transformer ini mampu memberikan hasil yang akurat, karena dapat memahami konteks bahasa Indonesia secara lebih mendalam dibandingkan metode tradisional.

Selain itu, penggunaan IndoBERT juga dinilai lebih mudah dipahami dalam implementasi, karena model ini sudah disediakan dalam bentuk *pre-trained model*

sehingga peneliti hanya perlu melakukan *fine-tuning* sesuai dengan kebutuhan dataset yang digunakan.

Beberapa studi terkait yang relevan antara lain:

Tabel 2. 1 Penelitian Terkait

NO	Penulis(Tahun)	Judul Penelitian	Hasil Penelitian
1	Gibran Hakim, Tirana Noor Fatyanosa, Agus WahyuWidodo (Hakim et al., 2023)	Analisis Sentimen Masyarakat terhadap Kereta Cepat Whoosh pada Platform X menggunakan IndoBERT	Penelitian ini menggunakan model IndoBERT untuk mengklasifikasikan sentimen publik terhadap proyek kereta cepat Jakarta-Bandung (Whoosh) di media sosial X. Hasilnya menunjukkan bahwa model ini efektif dalam menganalisis sentimen publik terkait infrastruktur transportasi besar.
2	MuhammadGhiyatsAl-Kadzim,Rasim Rasim, HerbertHerbert (Al-kadzim, 2024)	Analisis Perubahan Sentimen Publik di Media Sosial X terhadap Konflik	Penelitian ini menganalisis perubahan sentimen publik Indonesia di media sosial X terkait konflik

		Palestina-Israel Menggunakan Model IndoBERT	Palestina-Israel menggunakan model IndoBERT. Ditemukan bahwa fluktuasi sentimen sangat dipengaruhi oleh intensitas peristiwa kekerasan dan keputusan politik internasional
3	Indro Abri Oktariansyah, Fajri Rakhmat Umbara, Fatan Kasyidi (Oktariansyah et al., 2024)	Klasifikasi Sentimen Untuk Mengetahui Kecenderungan Politik Pengguna X Pada Calon Presiden Indonesia 2024 Menggunakan Metode IndoBert	Penelitian ini menggunakan model IndoBERT untuk menganalisis sentimen pengguna media sosial X terhadap calon presiden Indonesia 2024. Hasil penelitian menunjukkan bahwa augmentasi data dengan sinonim meningkatkan akurasi model dalam klasifikasi sentimen politik
4	Yessy Asri, Dwina Kuswardani, Widya Nita Suliyanti, Yosef Owen	Analisis Sentimen Pengguna terhadap Aplikasi	Penelitian ini menganalisis sentimen ulasan pengguna terhadap

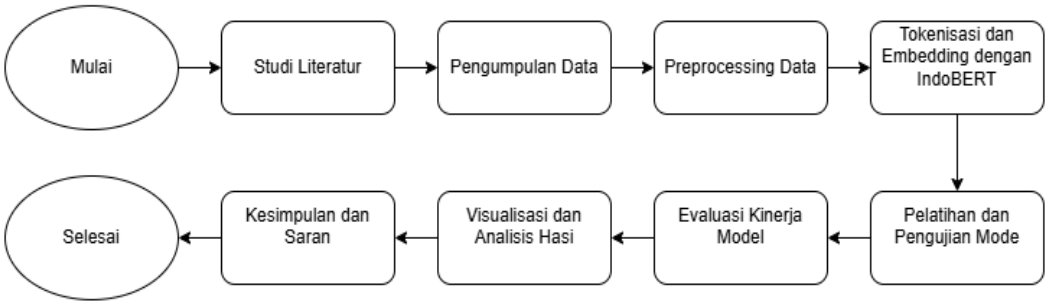
	Manullang, Atikah Rifdah Ansyari (Asri et al., 2025)	PLN Mobile Menggunakan IndoBERT dan Lexicon Bahasa Indonesia	aplikasi PLN Mobile dengan menggabungkan pendekatan berbasis leksikon bahasa Indonesia (INSET) dan model IndoBERT. Dari 1.000 ulasan yang dikumpulkan melalui web scraping, dilakukan klasifikasi sentimen positif, negatif, dan netral. Hasil analisis menunjukkan bahwa model IndoBERT mencapai akurasi sebesar 81%, menunjukkan efektivitasnya dalam memahami konteks bahasa Indonesia dalam ulasan pengguna.
5	Akbar Lucky Basuki, Bayu Rahayudi, Djoko Pramono (Basuki et al., 2025)	Analisis Sentimen Ulasan Aplikasi Ajaib Kripto menggunakan IndoBERT dan	Penelitian ini menggabungkan model IndoBERT dan metode Root Cause Analysis untuk menganalisis

		Metode Root Cause Analysis	sentimen ulasan pengguna aplikasi Ajaib Kripto. Ditemukan empat aspek utama keluhan pengguna, yaitu masalah transaksi, kinerja aplikasi, transaksi kripto, dan layanan aplikasi.

BAB III
METODOLOGI PENELITIAN

3.1 Alur Penelitian

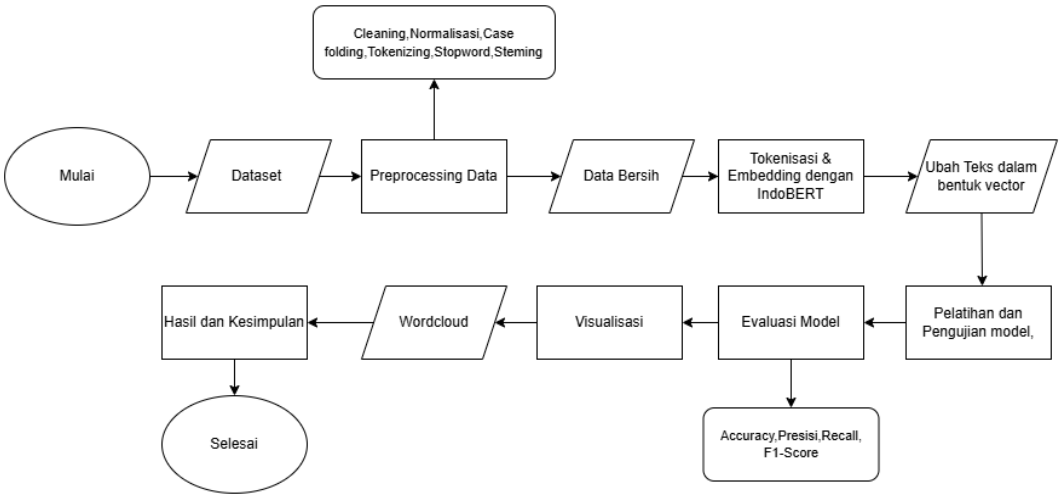
Alur Penelitian yang digambarkan pada gambar 3.1, digunakan sebagai tahapan-tahapan dalam penelitian ini guna mencapai tujuan dan menghindari penelitian keluar dari batasan masalah.



Gambar 3. 1 Alur Penelitian

3.2 Tahapan Implementasi

Tahapan implementasi dalam penelitian ini dirancang untuk menggambarkan alur teknis yang dilakukan dalam menganalisis sentimen masyarakat terhadap program Makan Bergizi Gratis menggunakan metode IndoBert Transformer. mulai dari pengumpulan data hingga evaluasi performa model. Selain proses utama tersebut, evaluasi model juga menjadi bagian penting dalam flowchart ini. Evaluasi dilakukan menggunakan confusion matrix untuk mengukur akurasi, presisi, recall dan f1-score dari hasil klasifikasi. Selanjutnya, dilakukan visualisasi data menggunakan WordCloud guna melihat kata-kata dominan dalam. Flowchart ini bertujuan memberikan gambaran menyeluruh mengenai jalannya implementasi analisis sentimen yang dilakukan dalam penelitian.



Gambar 3. 2 Flowchart Tahapan Implementasi

Penelitian ini diawali dengan tahap identifikasi masalah dan studi literatur, di mana peneliti merumuskan permasalahan utama yang berkaitan dengan opini publik terhadap program makan bergizi gratis dari pemerintah. Studi literatur dilakukan untuk menggali teori-teori yang relevan serta mengetahui pendekatan yang digunakan pada penelitian sebelumnya, khususnya dalam bidang analisis sentimen dan pemanfaatan model IndoBERT.

Setelah itu, dilakukan pengumpulan dan persiapan data, yang biasanya bersumber dari media sosial seperti X (Twitter) atau dataset terbuka seperti di Kaggle. Data ini kemudian diseleksi dan dipersiapkan untuk tahap selanjutnya. Pada tahap preprocessing data, dilakukan pembersihan teks melalui beberapa proses seperti cleaning (menghapus karakter tak perlu), case folding (mengubah huruf menjadi kecil), normalisasi (penyeragaman kata), tokenizing (memecah kalimat menjadi kata), stopword removal (menghapus kata tidak penting), dan stemming (mengubah kata ke bentuk dasar).

Langkah berikutnya adalah tokenisasi dan embedding menggunakan IndoBERT, yaitu mengubah teks ke dalam bentuk vektor yang dapat dipahami oleh

model transformer. Vektor-vektor ini kemudian digunakan dalam tahap pelatihan dan pengujian model, di mana data dilatih dan diuji untuk menghasilkan klasifikasi sentimen (positif, negatif, atau netral). Selanjutnya dilakukan evaluasi kinerja model dengan metrik seperti akurasi, presisi, recall, dan F1-score untuk mengukur seberapa baik model bekerja.

Hasil klasifikasi tersebut kemudian dianalisis secara visual melalui tahap visualisasi dan analisis hasil, seperti pembuatan grafik dan WordCloud yang membantu melihat persebaran opini dan kata-kata dominan dalam data. Akhirnya, penelitian ditutup dengan kesimpulan dan saran yang berisi temuan utama dari penelitian dan rekomendasi untuk pengembangan atau penelitian selanjutnya.

3.3 Metode Pengumpulan Data

Untuk penelitian ini diperlukan pengumpulan data agar dapat menghasilkan data yang lengkap yang dapat dijadikan acuan penelitian. Penulis penelitian ini memanfatkan sumber data yang berasal dari Kaggle serta studi terkait yang sejalan dengan analisis sentiment menggunakan metode Support Vector Machine.

3.3. 1 Studi Literatur

Studi pustaka adalah langkah penting dalam sebuah penelitian yang bertujuan untuk mengumpulkan, mengevaluasi, dan menganalisis berbagai sumber ilmiah yang terkait dengan topik yang diteliti. Biasanya, studi pustaka mencakup artikel jurnal, buku, laporan penelitian, dan sumber akademik lainnya yang memberikan dasar teori, konsep, serta hasil yang telah dirilis sebelumnya. Dengan melakukan studi pustaka, peneliti dapat memahami konteks masalah, menemukan celah dalam penelitian, serta memperkuat dasar teoritis dan metodologi yang akan

diterapkan. Selain itu, studi pustaka juga berkontribusi dalam merumuskan hipotesis dan mengenali metode terbaik untuk menjawab pertanyaan penelitian.

Melalui penelitian pustaka, peneliti dapat membandingkan berbagai perspektif serta menilai kelemahan dan kekuatan dari penelitian yang dilakukan sebelumnya. Hal ini membuka peluang untuk memperbarui atau mengembangkan konsep dalam penelitian yang sedang berlangsung. Studi pustaka tidak hanya memberikan pemahaman teoritis, tetapi juga membantu peneliti dalam memperkuat argumen dan mendukung dasar penelitian. Dengan kata lain, studi pustaka menjadi komponen penting dalam menyusun kerangka pemikiran serta memastikan bahwa penelitian memiliki relevansi dan kontribusi ilmiah yang berarti.

3.3. 2 Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari platform Kaggle. Kaggle merupakan sebuah komunitas data science dan machine learning yang menyediakan berbagai dataset yang dapat diakses secara terbuka oleh pengguna. Dataset yang digunakan dalam penelitian ini berjudul "makan gratis.csv", yang merupakan kumpulan data terkait sentimen masyarakat terhadap program pemerintah makan bergizi gratis yang di buat oleh engginer Altolyto Sitanggang (2024). Dengan total data sebanyak 656 tweet yang berisikan komentar Masyarakat mengenai program makan bergizi gratis pemerintahan presiden Prabowo subianto.

Dataset tersebut dipilih karena memiliki kesesuaian dengan tujuan penelitian yaitu melakukan analisis sentimen terhadap program makan bergizi gratis dengan menggunakan metode IndoBert Transformer. Data yang tersedia dalam dataset ini mencakup informasi berupa teks ulasan dan label sentimen (positif, negatif) yang sangat relevan untuk proses pengolahan data tekstual. Untuk

memastikan kualitas dan keakuratan data, dilakukan proses praproses data sebelum analisis lebih lanjut. Selain itu, sumber dataset juga disertai dengan lisensi yang memungkinkan penggunaan untuk tujuan penelitian.

Tabel 3. 1 Data Teks Tweet

No	Review	Label
1	@akunbarublek @ignasbowo Tpi tujuan awal dari makan siang gratis ini kan untuk perbaikan gizi dan pencegahan stunting? Pencegahan stunting ini jls g efektif krn prio nya di anak sekolah, lalu gmn kl misal anak itu ga suka lauknya/pnya alergi? Anak kecil kbnykn picky eater jd berkemungkinan besar g dimkn	N
2	BI Dukung Program Makan Siang Gratis Asalkan Tak Ganggu Stabilitas Keuangan https://t.co/OtJJgdUUOX	P
3	@RizkiDimasDwij @masgah_ @Ipaahu La iya dana bos aja dikorup, apalagi ada proyek makan siang gratis Sama aja kalau sdm sekolahnya sudah korup ya selamanya korup, korupnya tetep sama cuma beda proyek, kalau dulu korupnya proyek beli buku kalau skrng proyek beli susu	N
4	@OposisiCerdas Ya di gaji sama makan siang gratis lah	N
5	@RhendyW @marufins Sama kek stunting, solusinya makan siang gratis tanpa menyelesaikan akar permasalahannya (ditambah alurnya yang ga jelas)	N
6	@MrsTitins @ZackXChristy @akunbarublek @ignasbowo Gak baca programnya apa gimana ya? Targetnya aja makan	P

	siang gratis buat anak sekolah, yg putus sekolah ya otomatis gak dapet dong	
7	PKS: Kalau Dana BOS Dipakai Makan Siang Gratis, Terus Guru Honorer Mau Digaji Pakai Apa? https://t.co/1uStKOmGb1	N
8	https://t.co/LhpkGDxLR6 Bank Indonesia dukung Program Makan Siang Gratis	P
9	@KangManto123 makan siang gratis receh sampe harus pakai dana bos -_-	N
10	@YandraD31 Buat Gemoy Lover ndak sampai mikir ke situ bang.asal di imingi makan siang gratis aja udah bahagia koq.Ã°Å¸Ëœâ€™	N
	……..	

Data data diatas dijelaskan bahwa:

1. **Review** : Berisikan keluhan/ulasan Masyarakat di media sosial
2. **Sentimen** : Berisikan hasil antara positif dan negative yang diberi symbol N(negative) dan P(positif).

3.1 Pelabelan Encoding

Label encoding adalah teknik transformasi variabel kategorikal menjadi angka (biasanya integer) yang digunakan dalam tahap pra-pemrosesan data. Setiap kategori unik diberi angka berbeda untuk mempermudah proses input ke model

machine learning. Dalam penelitian oleh (Guntara & Astuti, 2025) label encoding dibandingkan dengan one-hot encoding pada algoritma K-Nearest Neighbor dan hasilnya menunjukkan bahwa meskipun label encoding lebih sederhana dan cepat, namun akurasi prediksi cenderung lebih rendah jika dibandingkan dengan one-hot encoding dalam data kategorikal tanpa urutan (Guntara & Astuti, 2025). Teknik label encoding sering kali digunakan dalam skenario klasifikasi biner, seperti dalam analisis sentimen, di mana label “positif” dan “negatif” diubah menjadi angka (misalnya, 1 dan 0) sebelum diproses oleh model.

Dalam konteks **analisis sentimen menggunakan IndoBERT Transformer**, label encoding memiliki peran penting dalam menyiapkan data sebelum masuk ke proses fine-tuning model. IndoBERT sebagai model Transformer berbasis arsitektur BERT hanya menerima input numerik dalam bentuk token ID dan label ID. Oleh karena itu, data sentimen seperti teks opini yang dilabeli "positif" atau "negatif" perlu dikonversi terlebih dahulu menjadi format numerik menggunakan label encoding agar dapat digunakan dalam proses pelatihan. Biasanya, label “positif” dikodekan menjadi 1 dan “negatif” menjadi 0. Proses ini penting untuk menghitung fungsi loss (seperti CrossEntropyLoss) yang digunakan selama pelatihan model IndoBERT, dan tanpa encoding yang tepat, model tidak akan bisa membedakan kelas target secara sistematis.

Dengan demikian, meskipun label encoding terlihat sebagai proses sederhana, dalam aplikasi analisis sentimen berbasis Transformer seperti IndoBERT, proses ini menjadi tahap krusial. Tidak hanya untuk memastikan kompatibilitas teknis dengan model, tetapi juga untuk menjaga interpretasi kelas target secara konsisten sepanjang eksperimen dan evaluasi performa model.

Dengan fungsi ini dapat merubah label yang sudah ada di dataset yang awalnya negative dan positif menjadi 0 untuk negative dan 1 untuk positif.

3.2 Preprocessing Data

Data preprocessing adalah proses mempersiapkan data mentah agar siap digunakan untuk analisis lebih lanjut atau model machine learning. Tujuannya adalah untuk membersihkan, mengubah, dan mengatur data sehingga menjadi lebih berkualitas, konsisten, dan sesuai dengan kebutuhan analisis atau.

3.4. 1 Case Folding

yaitu mengubah seluruh huruf dalam teks menjadi huruf kecil. Hal ini bertujuan untuk menyamakan format kata sehingga kata yang sama dengan kapitalisasi berbeda tidak dianggap berbeda. Misalnya, kata “Makan” dan “makan” akan diperlakukan sebagai satu kata yang sama.

Dengan melakukan case folding, sistem pengolahan bahasa alami (NLP) menjadi lebih konsisten dan efisien dalam mengelola data teks. Hal ini membantu meningkatkan akurasi dalam tugas-tugas seperti pencarian informasi, klasifikasi teks, dan analisis sentimen karena kata-kata yang sama diperlakukan secara seragam tanpa memperhatikan huruf kapital.

3.4. 2 Cleaning Data

Dalam analisis sentimen teks berbahasa Indonesia, tahap cleaning data atau pembersihan data sangat krusial untuk meningkatkan kualitas dan akurasi model. Menurut penelitian oleh (Sumanjaya et al., 2022), proses membersihkan data teks dari berbagai elemen yang tidak relevan atau mengganggu sebelum dilakukan analisis lebih lanjut. Dalam konteks pengolahan teks, proses ini mencakup penghapusan karakter khusus, angka, tanda baca, spasi berlebih, dan elemen lain

seperti URL atau emotikon yang tidak diperlukan. Tujuannya adalah menghasilkan data yang lebih bersih dan konsisten agar algoritma pengolahan bahasa alami dapat bekerja lebih efektif dan akurat.

Proses cleaning data sangat penting karena data mentah seringkali mengandung "noise" yang dapat mengganggu hasil analisis, seperti kesalahan ketik, simbol aneh, atau format yang tidak standar. Dengan membersihkan data, model dapat fokus pada informasi yang benar-benar bermakna dan mengurangi kemungkinan kesalahan dalam pemrosesan, seperti klasifikasi teks atau analisis sentimen. Cleaning data biasanya merupakan tahap awal dalam pipeline preprocessing data sebelum tahap lain seperti tokenizing, case folding, dan stopword removal dilakukan.

3.4. 3 Normalisasi Data

adalah proses mengubah kata-kata tidak baku, singkatan, kata slang, hingga kesalahan penulisan (typo) menjadi bentuk baku atau standar. Tujuan utama normalisasi adalah menyamakan bentuk kata sehingga analisis teks atau sentimen menjadi lebih akurat dan konsisten, normalisasi sangat penting sebelum tahapan seperti tokenisasi, stopword removal, atau stemming, karena jika tidak dinormalisasi, kata-kata yang sebenarnya bermakna sama akan dianggap berbeda oleh algoritma

3.3 Tokenisasi dan Embedding IndoBERT

Tokenisasi adalah tahap awal dalam menyiapkan teks untuk diproses oleh model IndoBERT, di mana kata atau sub-kata dipecah menjadi token. IndoBERT menggunakan tokenizer berbasis *WordPiece*–metode yang efektif untuk menangani kata-kata baru (*out-of-vocabulary*) dengan memecahnya menjadi bagian-bagian yang lebih kecil. Misalnya, kata "makanbergizi" dapat dipecah menjadi token

seperti “makan”, “##bergizi”. Hal ini memastikan setiap frasa dapat dijelaskan oleh model, meskipun belum ada dalam korpus latih sebelumnya.

Setelah tokenisasi, setiap token dikonversi ke dalam vektor *embedding*, yang merupakan representasi numerik berdimensi tinggi. Embedding ini bersifat kontekstual, artinya representasi kata berubah sesuai kalimatnya. Misalnya, kata “bank” dalam kalimat tentang keuangan memiliki embedding berbeda dibandingkan dalam konteks sungai. Embedding IndoBERT telah dipra-latih menggunakan korpus besar berbahasa Indonesia, sehingga mampu menghasilkan vektor yang reflektif terhadap relasi semantik antar kata.

Penelitian oleh (Imron et al., 2023) dalam "Deteksi Aspek Review E-Commerce Menggunakan IndoBERT Embedding dan CNN" menunjukkan embedding yang dihasilkan sangat kuat dalam menangkap konteks aspek review e-commerce, dengan akurasi mencapai 94,86 %. Ini menegaskan bahwa embedding IndoBERT dapat memperkaya model downstream—seperti CNN atau klasifier lainnya—untuk hasil yang lebih akurat.

Embedding ini kemudian digunakan sebagai input untuk tahap pelatihan model (fine-tuning atau feature-based), sehingga output akhir berupa klasifikasi sentimen (positif, negatif) berdasarkan konteks sebenarnya. Keseluruhan tahap tokenisasi dan embedding memastikan bahwa model transformator bekerja optimal dalam memahami dan mengklasifikasikan teks berbahasa Indonesia.

3.3 Konfigurasi Hyperparameter IndoBERT Transformer

Pada penelitian ini digunakan dataset berjumlah 656 data yang terdiri dari dua kelas sentimen, yaitu positif dan negatif. Data teks telah melalui tahap prapemrosesan seperti casefolding, cleaning, dan normalisasi sehingga siap

diproses lebih lanjut dengan model IndoBERT Transformer. Konfigurasi hyperparameter yang digunakan pada tahap pelatihan model disesuaikan dengan karakteristik dataset tersebut. Batch size ditetapkan sebesar 16, sehingga setiap epoch akan terbagi menjadi sekitar 41 iterasi. Nilai ini dianggap seimbang untuk ukuran dataset sedang dan mampu menjaga kestabilan proses pembelajaran. Learning rate yang digunakan adalah sebesar 2e-5, sesuai dengan rekomendasi penelitian sebelumnya pada model BERT, karena nilai ini cukup kecil untuk mencegah divergensi sekaligus memastikan pembaruan bobot berjalan stabil. Jumlah epoch ditetapkan sebanyak 3 kali agar model dapat belajar cukup dari dataset tanpa menimbulkan overfitting. Selain itu, weight decay ditetapkan sebesar 0.01 untuk mengurangi kompleksitas model dan meningkatkan kemampuan generalisasi, mengingat dataset yang digunakan relatif kecil. Dengan pengaturan hyperparameter ini, diharapkan IndoBERT mampu mempelajari pola sentimen dalam teks secara optimal dan menghasilkan prediksi yang akurat.

Dengan konfigurasi ini, diharapkan model IndoBERT Transformer dapat belajar secara optimal dari dataset yang digunakan. Pengaturan hyperparameter tersebut juga mengacu pada praktik umum dalam penelitian NLP berbasis BERT/IndoBERT, sehingga hasil yang diperoleh dapat dibandingkan dengan studi-studi terdahulu.

3.4 Pelatihan dan Evaluasi Model IndoBERT

Pada tahap ini, dilakukan proses pelatihan (fine-tuning) model IndoBERT dengan data yang telah melalui tahap pra-pemrosesan. Fine-tuning bertujuan untuk menyesuaikan parameter model IndoBERT agar dapat melakukan klasifikasi

sentimen berdasarkan data khusus yang digunakan dalam penelitian, yaitu opini masyarakat terhadap program makan bergizi gratis pemerintah.

Proses pelatihan dilakukan menggunakan platform Google Colab dengan memanfaatkan pustaka *Transformers* dari HuggingFace. Data dibagi menjadi dua bagian: data latih (Training data), data validasi (validation data) dengan perbandingan umum 85% dan 15%,Beberapa parameter pelatihan yang disesuaikan meliputi jumlah epoch, learning rate, dan batch size. Pada penelitian ini, digunakan IndoBERT versi indobert-base-p1, karena model ini telah terbukti efektif dalam tugas-tugas klasifikasi berbahasa Indonesia.

Setelah pelatihan selesai, model dievaluasi menggunakan data uji untuk mengukur kinerja model dalam mengklasifikasikan sentimen. Evaluasi dilakukan menggunakan metrik-metrik umum dalam klasifikasi teks, yaitu akurasi, presisi, recall, dan F1-score. Metrik ini memberikan gambaran menyeluruh tentang kualitas klasifikasi, terutama dalam mengatasi ketidakseimbangan label (unbalanced class). Seluruh hasil evaluasi akan ditampilkan dalam bentuk tabel dan grafik pada bab IV.

Penelitian ini mengacu pada pendekatan yang digunakan oleh(Anugerah Simanjuntak et al., 2024), yang berhasil menerapkan IndoBERT dengan teknik tuning hiperparameter untuk mendeteksi berita palsu dan memperoleh F1-score sebesar 91,56%. Oleh karena itu, metode evaluasi ini diharapkan memberikan gambaran menyeluruh terhadap kinerja model IndoBERT dalam konteks analisis sentimen masyarakat Indonesia.

3.5 Evaluasi Model

Dalam analisis sentimen, setelah data diproses dan diklasifikasikan, tahap evaluasi sangat penting untuk mengetahui sejauh mana model dapat memprediksi

sentimen dengan benar. Evaluasi dilakukan menggunakan confusion matrix, yang membandingkan hasil prediksi model dengan data yang sebenarnya. Tiga metrik utama yang sering digunakan untuk evaluasi adalah akurasi, presisi, dan recall.

1. Akurasi (Accuracy)

   Akurasi mengukur seberapa banyak prediksi yang benar dibandingkan dengan total keseluruhan data. Ini dihitung dengan rumus:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

   Di mana:

   a. **TP** = True Positive (prediksi positif yang benar)
   b. **TN** = True Negative (prediksi negatif yang benar)
   c. **FP** = False Positive (prediksi positif yang salah)
   d. **FN** = False Negative (prediksi negatif yang salah)

   Akurasi cocok digunakan ketika data memiliki distribusi yang seimbang antara kelas.

2. Presisi (Precision)

   Presisi menunjukkan seberapa akurat prediksi positif dari model. Artinya, dari semua data yang diprediksi sebagai positif, berapa persen yang benar-benar positif. Rumusnya:

$$\text{Presisi} = \frac{TP}{TP + FP}$$

   Presisi penting jika konsekuensi dari kesalahan prediksi positif cukup besar, seperti dalam analisis opini negatif terhadap kebijakan public.

3. Recall

Recall mengukur seberapa baik model dalam menemukan semua data yang benar-benar positif. Rumusnya:

$$\text{Presisi} = \frac{TP}{TP + FP}$$

Recall penting jika kita ingin memastikan semua sentimen negatif terdeteksi, meskipun ada kemungkinan memprediksi beberapa positif secara salah.

4. F1-Score

   Fokus utama dalam penelitian ini adalah pada metrik F1-score, karena F1-score memberikan penilaian yang seimbang antara presisi dan recall. Presisi mengukur berapa banyak prediksi positif yang benar, sedangkan recall mengukur berapa banyak data positif yang berhasil teridentifikasi. Rumus F1-score adalah:

$$\text{F1} - \text{score} = 2 \times \frac{Presisi \times Recall}{Presisi + Recall}$$

   F1-score sangat berguna dalam kondisi data tidak seimbang, di mana akurasi bisa menyesatkan. Misalnya, jika sebagian besar data bersentimen netral, model dapat terlihat memiliki akurasi tinggi hanya karena selalu memprediksi netral. Namun, F1-score akan menilai kinerja model secara adil berdasarkan keberhasilan dalam mengenali setiap kelas. Oleh karena itu, metrik ini menjadi ukuran utama untuk menilai efektivitas model IndoBERT dalam tugas analisis sentimen.

3.6 Visualisasi Data

WordCloud adalah teknik visualisasi data berbentuk kumpulan kata-kata yang disusun secara acak dengan ukuran font yang berbeda-beda. Ukuran kata mencerminkan frekuensi atau bobot kemunculan kata dalam kumpulan data. Dalam

konteks analisis sentimen, WordCloud digunakan untuk menampilkan kata-kata yang paling sering muncul dalam teks dengan sentimen tertentu (positif, negatif, atau netral)

Visualisasi WordCloud membantu peneliti atau pengguna untuk dengan cepat memahami kata-kata dominan dalam data, tanpa harus membaca semua teks secara manual. Misalnya, pada data sentimen negatif, kata-kata seperti “mahal”, “lambat”, atau “buruk” mungkin akan muncul dengan ukuran lebih besar karena sering disebut. WordCloud sangat efektif untuk eksplorasi awal data teks karena memberikan gambaran umum yang intuitif terhadap isi dokumen atau opini publik yang dianalisis.

WordCloud digunakan untuk menampilkan kata-kata dominan dalam masing-masing kelas sentimen, baik positif maupun negatif. Untuk sentimen positif, kata-kata yang sering muncul antara lain *"bagus"*, *"bermanfaat"*, *"tepat"*, *"peduli"*, *"membantu"*, dan *"sehat"*. Sedangkan untuk sentimen negatif, kata-kata yang dominan meliputi *"kurang"*, *"tidak tepat"*, *"tidak jelas"*, *"lelet"*, *"ribet"*, dan *"bermasalah"* Visualisasi ini memberikan gambaran umum mengenai persepsi masyarakat terhadap program yang dianalisis.

3.7 Perangkat Penelitian

Tabel 3. 2 Perangkat Penelitian

Hard ware	Laptop	Lenovo x260
Soft ware	OS	Windows
Tools	Kaggle	Goggle colab

3.8 Jadwal Penelitian

Tabel 3. 3 Jadwal Penelitian

No	Tahapan Penelitian	Bulan I	Bulan II	Bulan III	Bulan IV	Bulan V
1	Studi Literatur					
2	Penyusunan Proposal Penelitian					
3	Seminar Proposal					
4	Pengumpulan Data					
5	Pengolahan dan Analisis Data					
6	Implementasi dan Pengujian Model					
7	Penyusunan Laporan dan Dokumentasi					
8	Penyelesaian dan Sidang Akhir					

BAB IV
HASIL DAN PEMBAHASAN

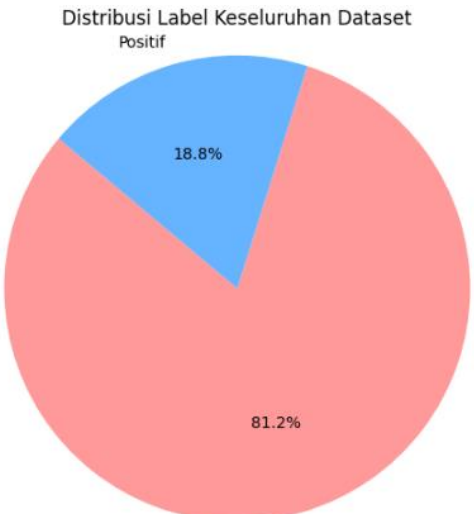
4.1 Gambaran Data Umum

Dataset yang digunakan dalam penelitian ini merupakan kumpulan data opini masyarakat dari media sosial terkait program pemerintah "Makan Gratis_valid.csv". Dataset ini telah melalui tahap pembersihan (cleaning) dan hanya memiliki dua kelas sentimen yaitu positif dan negatif. Total data yang digunakan sebanyak X data, yang telah dibagi menjadi:

Jumlah data negative : 533 atau sebesar 81.2%

Jumlah data positif : 123 atau sebesar 18.8%

Total data : 656 Data



Gambar 4. 1 Distribusi label

4.2 Hasil Preprocessing Data

Preprocessing merupakan tahap penting dalam analisis sentimen karena data mentah dari media sosial biasanya mengandung banyak "noise" seperti URL, emoji, tanda baca berlebih, serta penulisan kata yang tidak baku. Proses ini dilakukan agar

data lebih bersih dan siap diproses oleh model IndoBERT

4.2.1 Case Folding

Case folding merupakan salah satu tahap awal dalam preprocessing data teks yang bertujuan untuk menyeragamkan bentuk huruf. Proses ini dilakukan dengan cara mengubah semua huruf dalam teks menjadi huruf kecil (lowercase). Dengan demikian, kata yang sama tetapi berbeda penulisan huruf besar–kecilnya tidak dianggap sebagai kata yang berbeda oleh sistem, Sebagai contoh, kata *" Pencegahan"*, diubah menjadi *"pencegahan"* secara semantik memiliki makna yang sama.

Tabel 4. 1 Case Folding

Sentimen	Case Folding
@iweng01 @akunbarublek @ignasbowo Tpi tujuan awal dari makan siang gratis ini kan untuk perbaikan gizi dan pencegahan stunting? Pencegahan stunting ini jls g efektif krn prio nya di anak sekolah, lalu gmn kl misal anak itu ga suka lauknya/pnya alergi? Anak kecil kbnykn picky eater jd berkemungkinan besar g dimkn	@iweng01 @akunbarublek @ignasbowo tpi tujuan awal dari makan siang gratis ini kan untuk perbaikan gizi dan pencegahan stunting? pencegahan stunting ini jls g efektif krn prio nya di anak sekolah, lalu gmn kl misal anak itu ga suka lauknya/pnya alergi? anak kecil kbnykn picky eater jd berkemungkinan besar g dimkn

4.2.2 Cleaning Data

Cleaning data adalah proses membersihkan teks dari elemen-elemen yang tidak relevan. Pada data media sosial, teks sering kali mengandung URL, hashtag (#), mention (@username), angka, emotikon, tanda baca berlebihan, atau karakter khusus yang tidak memiliki makna dalam analisis sentiment

Tabel 4. 2 Cleaning Data

Case Folding	Cleaning Data
@iweng01 @akunbarublek @ignasbowo tpi tujuan awal dari makan siang gratis ini kan untuk perbaikan gizi dan pencegahan stunting? pencegahan stunting ini jls g efektif krn prio nya di anak sekolah, lalu gmn kl misal anak itu ga suka lauknya/pnya alergi? anak kecil kbnykn picky eater jd berkemungkinan besar g dimkn	tpi tujuan awal dari makan siang gratis ini kan untuk perbaikan gizi dan pencegahan stunting pencegahan stunting ini jls g efektif krn prio nya di anak sekolah lalu gmn kl misal anak itu ga suka lauknyapnya alergi anak kecil kbnykn picky eater jd berkemungkinan besar g dimkn

Sebagai contoh, teks *"@iweng01 @akunbarublek @ignasbowo tpi tujuan awal dari makan siang gratis ini kan untuk perbaikan gizi dan pencegahan stunting?"* setelah dibersihkan akan menjadi *"tpi tujuan awal dari makan siang gratis ini kan untuk perbaikan gizi dan pencegahan stunting pencegahan stunting"*. Dengan menghapus elemen-elemen tersebut, data menjadi lebih sederhana dan fokus pada isi pesan utama. Proses ini membantu mengurangi "noise" yang dapat

mengganggu kinerja model dan meningkatkan akurasi klasifikasi sentiment.

4.2.3 Normalisasi Data

Normalisasi data merupakan tahapan penting dalam preprocessing teks, terutama ketika data berasal dari media sosial yang penuh dengan singkatan, bahasa gaul, serta penulisan tidak baku. Proses ini bertujuan untuk menyeragamkan kata-kata agar sesuai dengan bentuk standar dalam bahasa Indonesia sehingga lebih mudah dipahami oleh model IndoBERT. Misalnya, kata "g" dinormalisasi menjadi "tidak", "yg" menjadi "yang", atau "krn" menjadi "karena". Dengan adanya normalisasi, kata-kata yang sebenarnya memiliki makna sama tidak lagi dianggap berbeda oleh model. Seperti pada tabel di bawah ini

Tabel 4. 3 Normalisasi Data

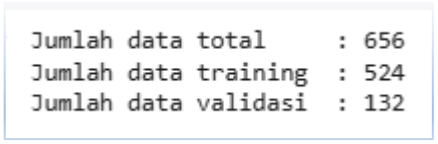
Cleaning Data	Normalisasi Data
tpi tujuan awal dari makan siang gratis ini kan untuk perbaikan gizi dan pencegahan stunting pencegahan stunting ini jls g efektif krn prio nya di anak sekolah lalu gmn kl misal anak itu ga suka lauknyapnya alergi anak kecil kbnykn picky eater jd berkemungkinan besar g dimkn	tapi tujuan awal dari makan siang gratis ini kan untuk perbaikan gizi dan pencegahan stunting pencegahan stunting ini jelas tidak efektif karena prio nya di anak sekolah lalu gmn kalau misal anak itu tidak suka lauknyapnya alergi anak kecil kebanyakan picky eater jadi berkemungkinan besar g dimkn

kalimat "tpi tujuan awal dari makan siang gratis ini kan untuk perbaikan gizi dan pencegahan stunting pencegahan stunting ini jls g efektif" setelah dinormalisasi menjadi

“tapi tujuan awal dari makan siang gratis ini kan untuk perbaikan gizi dan pencegahan stunting pencegahan stunting ini jelas tidak efektif”. Perubahan ini membuat representasi vektor yang dihasilkan IndoBERT menjadi lebih akurat karena kata-kata yang diproses telah sesuai dengan kaidah bahasa formal. Dengan demikian, normalisasi tidak hanya mengurangi keragaman kata yang tidak relevan, tetapi juga meningkatkan kualitas data sehingga model dapat lebih fokus memahami konteks sentimen yang terkandung dalam teks.

4.3 Split Data

Setelah data melalui tahap preprocessing, langkah berikutnya adalah melakukan pembagian data (split data) menjadi data latih dan data uji. Pembagian ini bertujuan agar model IndoBERT dapat dilatih menggunakan sebagian data, kemudian dievaluasi kinerjanya pada data yang belum pernah dilihat sebelumnya. Dengan demikian, performa model dapat diukur secara lebih objektif dan menghindari overfitting.



Gambar 4. 2 Jumlah Split data

Pada penelitian ini, dataset dibagi menggunakan perbandingan 80% untuk data latih (training set) dan 20% untuk data uji (validation/test set). Dari total 656 data, diperoleh sekitar 524 data latih dan 132 data uji. Distribusi kelas pada kedua bagian data tetap mempertahankan proporsi yang sama, yakni sentimen negatif lebih dominan daripada positif, sehingga model tetap menghadapi tantangan untuk menyeimbangkan prediksi. Proses split data ini menjadi pondasi penting sebelum

memasuki tahap tokenisasi dan embedding IndoBERT, karena kualitas pembagian data akan memengaruhi hasil evaluasi model secara keseluruhan.

4.4 Hasil Tokenisasi dan Embedding IndoBERT

Setelah data dibagi menjadi data latih dan data validasi, tahap berikutnya adalah melakukan tokenisasi menggunakan tokenizer bawaan dari IndoBERT. Tokenisasi berfungsi untuk memecah teks menjadi unit terkecil yang disebut token, sehingga model dapat memahami struktur kalimat secara lebih sistematis. IndoBERT menggunakan metode WordPiece Tokenizer, yang mampu menangani kata-kata baru (out-of-vocabulary) dengan cara memecahnya menjadi sub-kata.

Tabel 4. 4 Hasil Tokenisasi Indobert

Teks Asli	Tokenisasi	Token IDs
jelas akan memotong kepentingan sekolah yang lebih penting aneh program capres kenapa dibebankan pada dana bos cari dong tolak wacana program makan siang gratis pakai dana bos andreas hugo ke nadiem makarim merepotkan	'jelas', 'akan', 'memotong', 'kepentingan', 'sekolah', 'yang', 'lebih', 'penting', 'aneh', 'program', 'capres', 'kenapa', 'dibebankan', 'pada', 'dana', 'bos', 'cari', 'dong', 'tolak', 'wacana', 'program', 'makan', 'siang', 'gratis', 'pakai', 'dana', 'bos', 'andreas', 'hu', '##go', 'ke', 'nadi', '##em', 'makar', '##im'	1127, 150, 8409, 3041, 861, 34, 216, 906, 3526, 986, 15994, 2124, 18443, 126, 1869, 3235, 2203, 3339, 10348, 9000, 986, 521, 3346, 1243, 2468, 1869, 3235, 22719, 7925, 4294, 43, 16291, 25, 25719, 95, 18579

4.5 Hasil Pelatihan Model

Proses pelatihan model IndoBERT dilakukan menggunakan dataset yang telah melalui tahap preprocessing. Data dibagi menjadi 557 data latih (85%) dan 99 data validasi (15%). Pada tahap fine-tuning, model yang digunakan adalah indobert-base-p1 dengan konfigurasi hyperparameter sebagaimana ditunjukkan pada Tabel

Tabel 4. 5 Hyperparameter IndoBert

Hyperparameter	Skala
Batch size	16 dan 16
Learning rate	2e-5
Epoch	3
Weight Decay	0.01

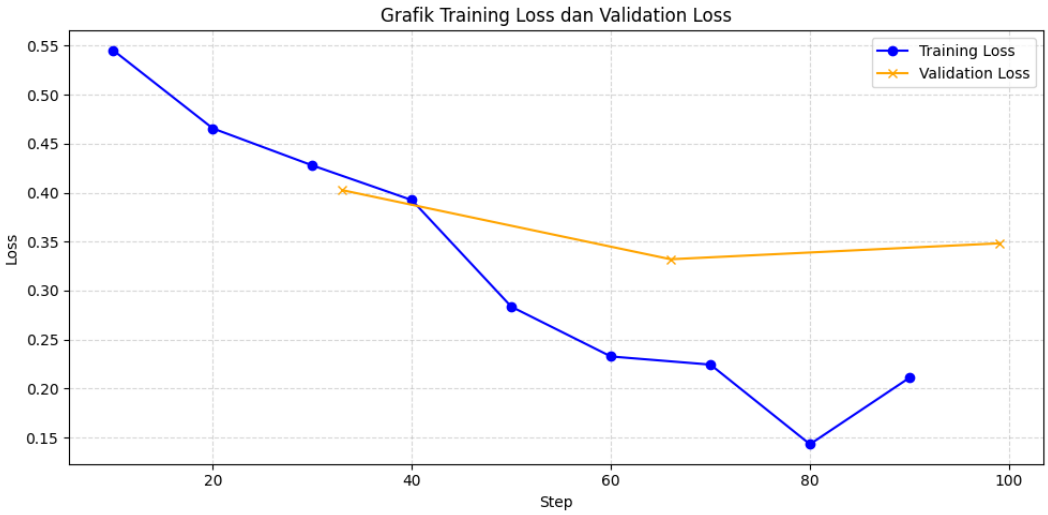
Pengaturan hyperparameter ini dipilih berdasarkan rekomendasi umum untuk fine-tuning model BERT, di mana *learning rate* **2e-5** dan jumlah epoch relatif kecil (**3 epoch**) sering digunakan untuk menjaga stabilitas pelatihan sekaligus menghindari *overfitting*. *Batch size* **16** dipilih untuk menyeimbangkan efisiensi komputasi dan stabilitas gradien, sementara *weight decay* sebesar **0.01** berfungsi sebagai regularisasi agar bobot model tidak terlalu besar.

Setelah proses pelatihan model IndoBERT dilakukan dengan konfigurasi hyperparameter pada Gambar 4.3, diperoleh nilai *training loss* dan *validation loss* pada setiap epoch. Nilai ini digunakan untuk memantau perkembangan proses

Epoch	Training Loss	Validation Loss
1	0.427800	0.402555
2	0.232700	0.331982
3	0.211000	0.348243

Gambar 4. 3 Hasil Epoch

pelatihan serta mengukur kemampuan model dalam melakukan generalisasi terhadap data validasi.



Gambar 4. 4 Grafik Training Loss

Perubahan nilai loss selama proses pelatihan divisualisasikan pada Gambar diatas dimana Pada epoch 1, nilai *training loss* sebesar 0,4278 dan *validation loss* sebesar 0,4025. Hal ini menunjukkan bahwa model masih berada pada tahap awal penyesuaian parameter dan belum optimal dalam mengenali pola data.

Pada epoch 2, *training loss* menurun signifikan menjadi 0,2327, sementara *validation loss* juga menurun menjadi 0,3319. Penurunan ini menandakan bahwa model semakin mampu belajar dari data dan generalisasi pada data validasi juga membaik,

Pada epoch 3, *training loss* kembali menurun cukup tajam menjadi 0,2110, sedangkan *validation loss* turun ke 0,3482. Perbedaan yang cukup stabil antara *training loss* dan *validation loss* mulai terlihat, namun tren penurunan pada *validation loss* tetap menunjukkan bahwa model tidak mengalami *overfitting* secara signifikan, Dengan demikian, proses pelatihan model IndoBERT dapat dikatakan stabil dan efektif, karena tidak ditemukan gejala *overfitting* yang berlebihan

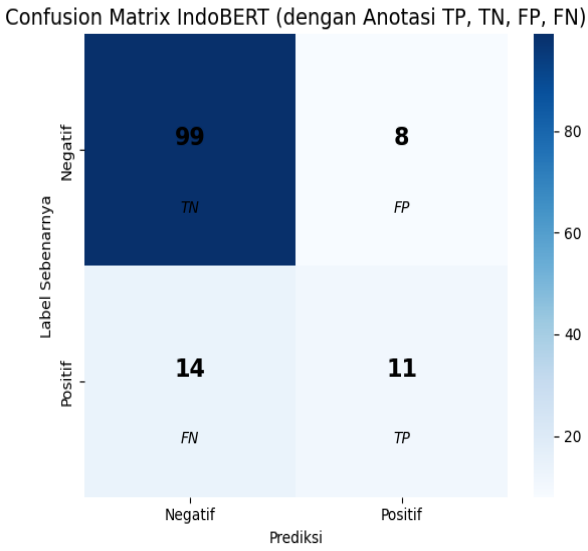
(dimana *validation loss* meningkat meskipun *training loss* terus menurun). Hasil ini menjadi dasar untuk melanjutkan tahap evaluasi performa model dengan menggunakan metrik akurasi, presisi, recall, dan F1-score.

4.6 Hasil Evaluasi Model

Evaluasi dilakukan untuk mengetahui sejauh mana model mampu melakukan klasifikasi sentimen dengan benar pada data uji, Hasil evaluasi ditunjukkan melalui metrik accuracy, precision, recall, dan f1-score. Selain itu, ditampilkan juga confusion matrix untuk memberikan gambaran distribusi prediksi model terhadap kelas positif maupun negatif.

4.6.1 Confusion Matrix

Confusion matrix pada Gambar 4.4 menunjukkan hasil perbandingan prediksi model IndoBERT dengan label sebenarnya pada data uji:



Gambar 4. 5 Confusion Matrix

Dari confusion matrix diperoleh:

1. True Negative (TN) = 99 → data negatif yang berhasil diprediksi benar.
2. False Positive (FP) = 8 → data negatif yang salah diprediksi positif.
3. False Negative (FN) = 14 → data positif yang salah diprediksi negatif.

4. True Positive (TP) = 11 → data positif yang berhasil diprediksi benar.

Hasil ini menunjukkan bahwa model relatif lebih baik dalam mengenali sentimen negatif (TN tinggi) dibandingkan sentimen positif (TP lebih sedikit).

4.6.2 Classification Report

Dari confusion matrix tersebut, diperoleh metrik evaluasi yang dihitung menggunakan *precision*, *recall*, *F1-score*, dan *accuracy*. Metrik ini penting karena memberikan gambaran menyeluruh mengenai kemampuan model, bukan hanya dari sisi akurasi, tetapi juga bagaimana model dalam mengklasifikasikan setiap kelas secara seimbang

Dari confusion matrix tersebut, diperoleh metrik evaluasi sebagai berikut:

	precision	recall	f1-score	support
negatif	0.88	0.93	0.90	107
positif	0.58	0.44	0.50	25
accuracy			0.83	132
macro avg	0.73	0.68	0.70	132
weighted avg	0.82	0.83	0.82	132

Gambar 4. 6 Metrik evaluasi

Penjelasan hasil:

1. Pada kelas Negatif, model sangat baik dengan *precision* 0.88 dan *recall* 0.93. Hal ini menandakan mayoritas data negatif berhasil dikenali dengan benar. Nilai recall yang tinggi menandakan bahwa model memiliki kemampuan yang baik dalam mengenali opini negatif yang terdapat pada data uji.
2. Pada kelas Positif, *precision* 0.58 dan *recall* 0.44, dengan *fi-score* sebsar 0.50 cukup baik, meskipun masih ada beberapa data positif yang salah diklasifikasikan sebagai negatif.
3. Nilai akurasi sebesar 0.83 (83%) menunjukkan bahwa model memiliki

kemampuan yang cukup baik dalam mengklasifikasikan teks ke dalam dua kategori sentimen.

Hasil ini mengindikasikan bahwa model stabil dan mampu mengenali pola umum dari teks berbahasa Indonesia, khususnya dalam konteks opini terhadap program "Makan Bergizi Gratis".

4. Macro average (0.70) Nilai macro average sebesar 0.70 menunjukkan performa rata-rata dari kedua kelas (positif dan negatif).Karena data negatif lebih dominan, maka performa pada kelas positif cenderung lebih rendah sehingga menurunkan nilai macro average.Hal ini menunjukkan adanya ketidakseimbangan distribusi kelas yang dapat memengaruhi hasil keseluruhan.
5. Weighted average sebesar (0.82) menunjukkan performa keseluruhan yang tetap tinggi karena bobot terbesar berasal dari kelas negatif yang mendominasi dataset.

   Meskipun demikian, model tetap memperlihatkan hasil yang konsisten dengan kinerja stabil pada data uji.

4.6.3 Analisis

Hasil evaluasi menunjukkan bahwa:

1. Model seimbang → Tidak hanya mendominasi pada kelas negatif, tetapi juga mampu mengenali cukup baik kelas positif (walaupun lebih rendah dari negatif).
2. Kesalahan prediksi → Masih ada 8 data negatif yang salah diprediksi sebagai positif (FP) dan 14 data positif yang salah diprediksi sebagai negatif (FN).

3. Tingkat generalisasi baik → Dengan akurasi 83% dan F1-score di atas 0.60 untuk kedua kelas, model dapat dikatakan efektif dalam melakukan analisis.

4.7 Visualisasi Data Hasil Analisis

Selain menampilkan hasil evaluasi berupa confusion matrix dan classification report, analisis juga diperkuat dengan visualisasi data. Visualisasi ini bertujuan untuk memberikan gambaran yang lebih intuitif mengenai hasil prediksi model IndoBERT terhadap data uji.

4.71 Wordcloud

WordCloud digunakan untuk menampilkan kata-kata yang paling sering muncul pada setiap label sentimen. Semakin besar ukuran kata, semakin sering kata tersebut muncul dalam data.

Gambar 4. 7 Wordcloud Positif

menunjukkan kata-kata dominan seperti *siang*, *gratis*, *program*, *anak-anak*, *penting*, dan *buat*. Kata *siang* dan *gratis* sering muncul karena konteks program yang memang berupa *makan siang gratis*. Kata *program* dan *penting* menggambarkan apresiasi masyarakat yang menilai kebijakan ini bermanfaat, khususnya bagi *anak-anak* dan dunia pendidikan. Kehadiran kata *buat*, *pak*, dan *banget* memperlihatkan gaya bahasa sehari-hari masyarakat dalam memberikan dukungan. Secara keseluruhan, kata-kata ini menekankan bahwa sentimen positif

lebih banyak berfokus pada manfaat program dan dampak langsung bagi siswa dan orang tua.

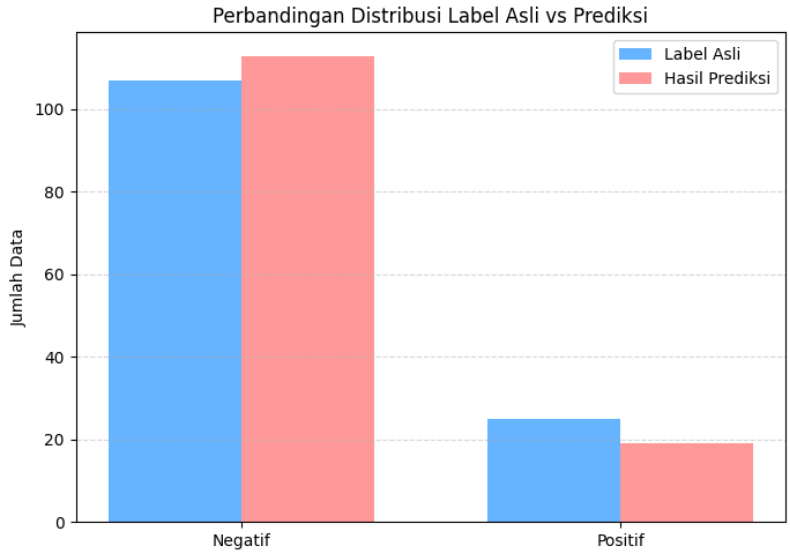
Gambar 4. 8 Wordcloud negatif

memperlihatkan kata-kata seperti *tidak*, *jadi*, *saja*, *malah*, *penting*, *program*, dan *pendidikan*. Kata *tidak* dan *malah* banyak muncul pada kalimat yang bernada kritik, misalnya “tidak perlu”, “malah membebani”, atau “tidak penting”. Kata *pendidikan* dan *program* menunjukkan adanya perdebatan mengenai prioritas kebijakan, di mana sebagian masyarakat menilai anggaran makan gratis sebaiknya dialihkan ke sektor pendidikan lain. Kehadiran kata *bikin*, *pilih*, dan *rakyat* memperlihatkan nuansa kritik terhadap implementasi maupun kepentingan politik di balik program ini. Dengan demikian, sentimen negatif lebih banyak menyoroti keterbatasan, kendala, dan keraguan terhadap efektivitas program.

4.72 Distribusi Label Prediksi

Selain WordCloud, analisis hasil model IndoBERT juga divisualisasikan melalui distribusi jumlah label prediksi. Distribusi ini penting karena tidak hanya memperlihatkan seberapa banyak data uji yang diprediksi sebagai **positif** dan **negatif**, tetapi juga memperlihatkan sejauh mana hasil prediksi model mendekati distribusi label asli. Dengan kata lain, distribusi ini menjadi indikator tambahan

untuk menilai kemampuan model dalam melakukan generalisasi terhadap data uji.



Gambar 4. 9 Perbandingan Distribusi Label Asli vs Prediksi

Penjelasan:

1. Pada kelas negatif, label asli berjumlah sekitar 107 data, sedangkan hasil prediksi model mencapai sekitar 112 data.Artinya, model IndoBERT sedikit melebihkan jumlah prediksi pada kelas negatif.Hal ini menunjukkan bahwa model memiliki kecenderungan lebih kuat dalam mendeteksi sentimen negatif, yang juga selaras dengan hasil confusion matrix di mana *True Negative (TN)* memiliki jumlah tertinggi.Meskipun demikian, perbedaan antara label asli dan hasil prediksi tidak terlalu signifikan, yang berarti model mampu memprediksi data negatif dengan akurasi yang baik.
2. Pada kelas positif, label asli berjumlah sekitar 25 data, sementara hasil prediksi model menurun menjadi sekitar 19 data.Kondisi ini menunjukkan bahwa sebagian data positif masih salah diklasifikasikan sebagai negatif (*False Negative*).Hal ini dapat terjadi karena variasi kalimat positif dalam bahasa Indonesia lebih beragam dan sering kali bersifat implisit, sehingga model masih kesulitan mengenalinya secara konsisten.

3. Secara keseluruhan, distribusi hasil prediksi model IndoBERT menunjukkan pola yang cukup mendekati distribusi label asli.Model mampu merepresentasikan kecenderungan data dengan baik, meskipun performa pengenalan pada kelas positif masih sedikit di bawah kelas negatif.Pola ini juga konsisten dengan hasil confusion matrix dan classification report, di mana performa model lebih dominan pada kelas negatif.Meskipun demikian, perbedaan jumlah prediksi antar kelas tetap tergolong kecil, sehingga model dapat dikatakan memiliki kemampuan generalisasi yang cukup baik dalam membedakan dua label sentimen — positif dan negatif.

Distribusi ini konsisten dengan hasil pada confusion matrix dan classification report, di mana model lebih unggul dalam mengenali kelas negatif dibandingkan kelas positif. Namun demikian, meskipun prediksi pada kelas positif sedikit lebih rendah, jumlahnya tetap seimbang dengan distribusi asli. Hal ini menandakan bahwa IndoBERT tetap memiliki kemampuan generalisasi yang baik dalam menangani dua label sentimen sekaligus.

4.73 Contoh Hasil Prediksi Model

Hasil prediksi model menunjukkan bagaimana IndoBERT Transformer menafsirkan setiap teks dalam dataset uji dan memprediksi label sentimennya, misalnya negatif atau positif. Prediksi ini merupakan langkah penting karena menghubungkan proses pelatihan model dengan penilaian kinerja aktual pada data yang belum pernah dilihat sebelumnya. Dengan kata lain, hasil prediksi ini merepresentasikan kemampuan model dalam memahami konteks dan makna teks bahasa Indonesia, serta menafsirkan sentimen yang terkandung di dalamnya.

7	tenang ada makan siang gratis ðÿªðÿª	positif	positif	✅ Benar
8	kan yang menikmati makan siang gratis cuma anak doang yg dewasa cari sendiri makanan sana ðÿ sdh di bilang masih saja tetap all in all in gass ok ðÿâðÿ di kasih calon yg mementingkan pendidikan gratis eh malah pilih makan gratis	negatif	negatif	✅ Benar
9	senang deh program makan siang gratis sudah dibuat testimoni ya dulu makasih pak saya terharu pak sangat mengedepankan makan ketimbang internet paksatu langkah dan trobosan kedepan semoga kelak kalian bs buat bangga bangsa adik manis	positif	positif	✅ Benar
10	yang penting makan siang gratis pak makan lebih butuh drpd pendidikan	positif	negatif	❌ Salah
11	gilak ini mah udah dipermudah wkwkwk malah milih makan gratis mana siang pulak bukan tiap hari wkwkwo	negatif	negatif	✅ Benar
12	kondisi seperti ini msh banyak dipelosok negeri dan ribuan bangunan sekolah yg tidak layak utk digunakan kok mau bikin program makan siang gratis dgn anggaran t lebihtahun otak dimanaðÿƒ	negatif	negatif	✅ Benar
13	katanya ke rakyat jangan pelit kok ngga setuju ngasih makan rakyat jk soal makan siang gratis bebani apbn siapa yang bayar kita semua	positif	negatif	❌ Salah
14	rakyat indonesia kena prank habis makan bansos dan makan siang gratis disuruh makan tainya sendiri ðÿ	negatif	negatif	✅ Benar

Gambar 4. 10 Hasil Prediksi Model

Interpretasi hasil prediksi model IndoBERT Transformer menunjukkan sejauh mana model mampu mengenali dan mengklasifikasikan sentimen teks dengan benar. Dari tabel hasil prediksi, dapat dilihat berapa banyak teks yang dikategorikan sesuai dengan label asli dan berapa banyak yang salah diprediksi. Analisis ini penting karena memungkinkan peneliti menilai kinerja model secara lebih mendetail, tidak hanya berdasarkan metrik numerik seperti akurasi, precision, recall, atau F1-score, tetapi juga berdasarkan kasus nyata di setiap teks. Selain itu, interpretasi hasil ini membantu mengidentifikasi pola kesalahan, misalnya teks yang ambigu, mengandung ironi, atau kata-kata yang jarang muncul dalam dataset sehingga model mengalami kesulitan. Data hasil prediksi ini juga menjadi dasar untuk visualisasi seperti confusion matrix, yang mempermudah pemahaman terhadap distribusi prediksi benar dan salah, serta word cloud yang menggambarkan kata-kata yang sering muncul pada tiap kategori sentimen.

4.8 Pembahasan Hasil Penelitian

Berdasarkan hasil pelatihan dan evaluasi, model IndoBERT terbukti mampu melakukan klasifikasi sentimen terhadap program makan bergizi gratis dengan akurasi yang cukup tinggi, yaitu 83%. Hasil ini menunjukkan bahwa pendekatan

berbasis *transformer* khususnya IndoBERT memiliki kinerja yang baik dalam memahami bahasa Indonesia, terutama dalam konteks teks media sosial dan opini publik.

Dari confusion matrix dan classification report, terlihat bahwa model lebih unggul dalam mengenali sentimen negatif dibandingkan positif. Hal ini disebabkan oleh ketidakseimbangan data dalam dataset, di mana jumlah data negatif lebih banyak dibanding positif. Dampaknya, model lebih terlatih untuk mengenali pola kalimat negatif sehingga *recall* kelas positif relatif lebih rendah. Meskipun demikian, nilai *precision* pada kelas positif tetap cukup baik, yang menunjukkan bahwa ketika model memprediksi positif, prediksi tersebut umumnya benar.

Visualisasi data seperti WordCloud dan distribusi label prediksi semakin memperkuat temuan ini. WordCloud memperlihatkan kata-kata dominan pada masing-masing kelas sentimen, sedangkan distribusi label prediksi menunjukkan kecenderungan model dalam lebih sering mengklasifikasikan kalimat sebagai negatif. Contoh hasil prediksi model juga memperlihatkan bahwa IndoBERT mampu mengenali banyak pola bahasa dengan tepat, meskipun masih terdapat beberapa kesalahan pada kalimat dengan makna implisit atau ambigu.

Jika dibandingkan dengan penelitian terdahulu yang menggunakan metode *machine learning* tradisional seperti Naive Bayes dan Support Vector Machine (SVM), penggunaan IndoBERT pada penelitian ini memberikan hasil yang lebih baik. Hal ini sesuai dengan temuan beberapa penelitian sebelumnya yang menyebutkan bahwa model berbasis *transformer* lebih unggul dalam menangkap konteks semantik bahasa dibanding metode klasik.

Namun demikian, penelitian ini juga memiliki keterbatasan. Pertama, dataset

yang digunakan masih relatif kecil dan tidak seimbang, sehingga berdampak pada performa model terutama pada kelas minoritas. Kedua, model hanya diuji pada satu jenis domain teks tertentu, yaitu opini publik tentang program pemerintah, sehingga generalisasi ke domain lain masih perlu dibuktikan.

Dengan demikian, hasil penelitian ini dapat menjadi dasar untuk pengembangan penelitian selanjutnya, misalnya dengan memperbesar ukuran dataset, melakukan *data augmentation* untuk menyeimbangkan kelas, atau melakukan *hyperparameter tuning* lebih lanjut. Selain itu, penerapan model ini juga berpotensi untuk digunakan sebagai alat bantu dalam analisis opini publik secara real-time terhadap kebijakan pemerintah.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan mengenai analisis sentimen program makan bergizi gratis menggunakan metode IndoBERT Transformer, diperoleh beberapa kesimpulan sebagai berikut:

1. Model IndoBERT mampu melakukan klasifikasi sentimen teks berbahasa Indonesia dengan cukup baik. Hal ini ditunjukkan dengan akurasi sebesar 83% pada data uji.
2. Model lebih unggul dalam mengenali sentimen negatif dengan nilai *precision* sebesar 0.88, *recall* 0.93, dan *F1-score* 0.90. Sementara itu, pada sentimen positif performanya masih lebih rendah dengan *precision* 0.58, *recall* 0.44, dan *F1-score* 0.50.
3. Faktor ketidakseimbangan data berpengaruh terhadap performa model, di mana jumlah data negatif yang lebih dominan membuat model cenderung bias terhadap kelas negatif.
4. Visualisasi data melalui WordCloud, distribusi label, dan contoh prediksi model memperlihatkan pola kata yang dominan dalam masing-masing kelas serta menunjukkan bahwa IndoBERT dapat mengenali banyak kalimat dengan tepat, meskipun masih terdapat kesalahan pada teks yang ambigu.
5. Dibandingkan metode *machine learning* tradisional, penggunaan IndoBERT terbukti memberikan hasil yang lebih baik dalam menangkap konteks semantik bahasa Indonesia.

5.2 Saran

Penelitian ini masih memiliki keterbatasan, sehingga diperlukan pengembangan lebih lanjut. Adapun saran yang dapat diberikan antara lain:

1. Perluasan dataset dengan jumlah data yang lebih besar dan seimbang antara kelas positif dan negatif agar model dapat melakukan generalisasi lebih baik.
2. Eksperimen hyperparameter tuning lebih lanjut, misalnya dengan variasi *learning rate*, jumlah epoch, atau ukuran *batch size*, untuk menemukan konfigurasi optimal.
3. Menguji model pada domain teks yang berbeda (misalnya berita, komentar produk, atau opini politik) untuk mengetahui kemampuan generalisasi IndoBERT di luar konteks program pemerintah.
4. Mengintegrasikan model ini ke dalam sistem analisis opini publik secara web sehingga dapat digunakan sebagai alat bantu evaluasi kebijakan pemerintah.

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.

Alfarobby, A. N., & Irawan, H. (2024). Analisis Sentimen Kepuasan Konsumen Pengguna Transportasi Online Pada Ulasan Google Playstore Menggunakan Indobert Dan Topic Modeling (Studi kasus: Gojek dan Grab). *E-Proceeding of Management*, *11*(1), 72.

Amien, M., Frendi Gunawan, G., & Kunci, K. (2024). ELANG: Journal of Interdisciplinary Research BERT dan Bahasa Indonesia: Studi tentang Efektivitas Model NLP Berbasis Transformer. *ELANG: Journal of Interdisciplinary Research*. https://jurnal.stiki.ac.id/elang/article/view/1152

Andin, A., Risti, D., Latifah, I., Panuntun, M., Nur, M., Selviani, R., Yogyakarta, U. P., Bantul, K., & Istimewa, P. D. (2025). *Penerapan Nilai Pancasila Melalui Program Makan Bergizi Gratis*. *3*(1), 370–383.

Anugerah Simanjuntak, Rosni Lumbantoruan, Kartika Sianipar, Rut Gultom, Mario Simaremare, Samuel Situmeang, & Erwin Panggabean. (2024). Research and Analysis of IndoBERT Hyperparameter Tuning in Fake News Detection. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *13*(1), 60–67. https://doi.org/10.22146/jnteti.v13i1.8532

Apriyadi, C. (2025). *Sentiment Analysis of Cyber Attacks in Bank Syariah Indonesia Using SVM and Indobert Method*. *6*(2), 819–838.

Arifadilah, F. (2023). *Perbandingan hyperparameter optimization population based training, random search, bayesian optimization pada analisis sentimen*

radikalisme. Fakultas Sains dan Teknologi UIN Syarif HIdayatullah Jakarta.

Asri, Y., Kuswardani, D., Suliyanti, W. N., & Manullang, Y. O. (2025). *Sentiment analysis based on Indonesian language lexicon and IndoBERT on user reviews PLN mobile application*. *38*(1), 677–688. https://doi.org/10.11591/ijeecs.v38.i1.pp677-688

Basuki, A. L., Rahayudi, B., Pramono, D., Studi, P., Informasi, S., Komputer, F. I., Brawijaya, U., Aplikasi, K., & Kripto, T. (2025). *ANALISIS SENTIMEN ULASAN APLIKASI AJAIB KRIPTO*. *9*(4), 1–9.

Gunawan, I. P. (2021). Statistika Deskriptif menggunakan R pada Google Colab untuk Ilmu Komputer. *Universitas Bakrie*, 6.

Guntara, M., & Astuti, F. D. (2025). Komparasi Kinerja Label-Encoding dengan One-Hot-Encoding pada Algoritma K-Nearest Neighbor menggunakan Himpunan Data Campuran. *JIKO (Jurnal Informatika Dan Komputer)*, *9*(2), 352. https://doi.org/10.26798/jiko.v9i2.1605

Hakim, G., Fatyanosa, T. N., & Widodo, A. W. (2023). *Analisis Sentimen Masyarakat terhadap Kereta Cepat Whoosh pada Platform X menggunakan IndoBERT*. *1*(1), 1–10.

Hendraswari, D. E. (2023). Studi Ekologi : Determinan Kejadian Gizi Buruk Di Indonesia Tahun 2021. *J-KESMAS: Jurnal Kesehatan Masyarakat*, *9*(1), 84. https://doi.org/10.35329/jkesmas.v9i1.3899

Hidayat, M. N., & Pramudita, R. (2024). Analisis Sentimen Terhadap Pembelajaran Secara Daring Pasca Pandemi Covid-19 Menggunakan Metode IndoBERT. *INFORMATION MANAGEMENT FOR EDUCATORS AND PROFESSIONALS : Journal of Information Management*, *8*(2), 161.

https://doi.org/10.51211/imbi.v8i2.2719

Imron, S., Setiawan, E. I., & Santoso, J. (2023). Deteksi Aspek Review E-Commerce Menggunakan IndoBERT Embedding dan CNN. *Journal of Intelligent System and Computation*, *5*(1), 10–16. https://doi.org/10.52985/insyst.v5i1.267

Jayadianti, H., Kaswidjanti, W., Utomo, A. T., Saifullah, S., Dwiyanto, F. A., & Drezewski, R. (2022). Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN. *ILKOM Jurnal Ilmiah*, *14*(3), 348–354. https://doi.org/10.33096/ilkom.v14i3.1505.348-354

Maulana, H., & Al-Khowarizmi, A.-K. (2021). Analysis of the Effectiveness of Online Learning Using Eda Data Science and Machine Learning. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *6*(1), 222–231.

Mubarok, M. I., & Tanjung, M. A. P. (2024). IMPLEMENTASI NATURAL LANGUAGE PROCESSING DALAM PERANCANGAN APLIKASI CHATBOT PADA FIKTI UMSU. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(6), 11992–12001.

Musfiroh, D., Khaira, U., Utomo, P. E. P., & Suratno, T. (2021). Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *1*(1), 24–33. https://doi.org/10.57152/malcom.v1i1.20

Mustaqim, H., Dzaky, M., Bima, P., Irfansyah, F., Ilham, S. N., Hasna, S., Ilmu, F., Dan, S., Politik, I., Jakarta, U. M., Dahlan, K. H. A., Tim, K. C., Selatan, K. T., Hukum, I., Hukum, F., Jakarta, U. M., Dahlan, J. K. H. A., Tim, K. C., Selatan, K. T., … Jakarta, U. M. (2024). *Penanggulangan Masalah Gizi Buruk*

Balita dengan Edukasi Pencegahan Stunting di Posyandu Pondok Cabe Udik. *November*, 1–10.

Nasution, M. D., Al-Khowarizmi, A.-K., & Maulana, H. (2023). Optimization of Faster R-CNN to Detect SNI Masks at Mandatory Mask Doors. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, *6*(2), 602–611.

Nugroho, R., Azka, N., Sayudha, W., & Graha, R. (2025). *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi) Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google*. *9*(June), 495–505.

Oktariansyah, I. A., Umbara, F. R., & Kasyidi, F. (2024). Klasifikasi Sentimen Untuk Mengetahui Kecenderungan Politik Pengguna X Pada Calon Presiden Indonesia 2024 Menggunakan Metode IndoBert. *Building of Informatics, Technology and Science (BITS)*, *6*(2), 636–648. https://ejurnal.seminar-id.com/index.php/bits/article/view/5435

Rahman, R. D., Setiawan, N. Y., & Bachtiar, F. A. (2025). *Analisis Sentimen Pengguna Aplikasi Mobile Berbasis Review Pada Platform Blibli Menggunakan Metode Bidirectional Encoder Representations from Transformers (BERT)*. *9*(4), 1–9.

Report, F. (2024). *Efek Pengganda Program Makan Bergizi Gratis*.

Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *1*(12), 1725–1732. http://j-

ptiik.ub.ac.id/index.php/j-ptiik/article/view/628

Sumanjaya, A. A. A., Indriati, & Ridok, A. (2022). Analisis Sentimen Data Tweets terhadap Penanganan Covid-19 di Indonesia menggunakan Metode Naïve Bayes dan Pemilihan Kata Bersentimen menggunakan Lexicon Based. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *6*(4), 1865–1872. http://j-ptiik.ub.ac.id

LAMPIRAN

1. Sk- Pembimbing

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
https://fikti.umsu.ac.id fikti@umsu.ac.id umsumedan umsumedan umsumedan umsumedan

PENETAPAN DOSEN PEMBIMBING
PROPOSAL/SKRIPSI MAHASISWA
NOMOR : 45/II.3-AU/UMSU-09/F/2025

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

Program Studi	**: Teknologi Informasi**
Pada tanggal	**: 03 Januari 2025**

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama	**: Handhalla Akasyah Alkautsar Panjaitan**
NPM	**: 2109020012**
Semester	**: VII (Tujuh)**
Program studi	**: Teknologi Informasi**
Judul Proposal / Skripsi	**: Sistem Pemantauan Konsumsi Energi Listrik Berbasis IoT dengan Metode Clustering K-Means dan Support Vector Machine (SVM) untuk Prediksi Beban**
Dosen Pembimbing	**: Mahardika Abdi Prawira, S.Kom.,M.Kom.**

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi** dinyatakan " **BATAL** " bila tidak selesai sebelum Masa Kadaluarsa tanggal : **03 Januari 2026**
4. Revisi judul……

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

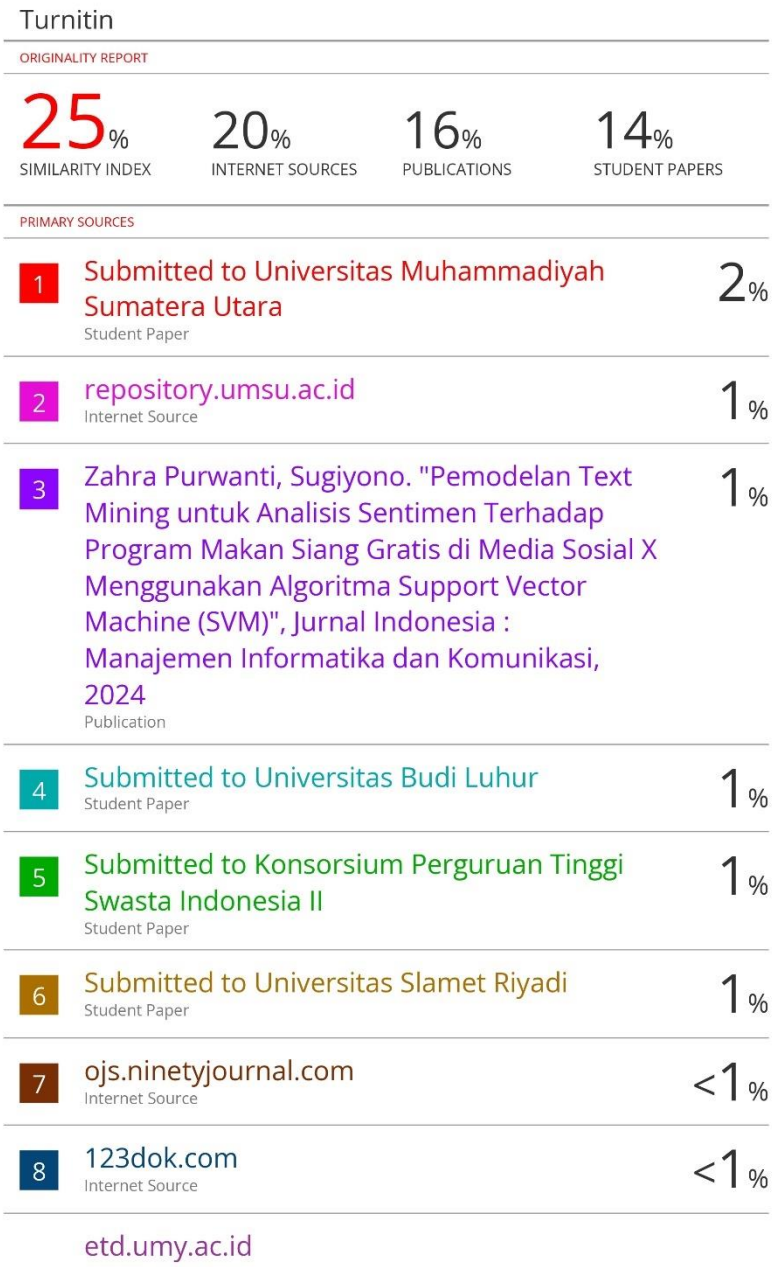
Ditetapkan di : Medan
Pada Tanggal : 03 Rajab 1446 H
03 Januari 2025 M

a.n.Dekan
Wakil Dekan I

Hadrik Maulana, S.T.,M.Kom.
NIDN : 0121119102

Cc. File

2. Hasil Turnitin



Turnitin

ORIGINALITY REPORT

25%	20%	16%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Muhammadiyah Sumatera Utara Student Paper	2%
2	repository.umsu.ac.id Internet Source	1%
3	Zahra Purwanti, Sugiyono. "Pemodelan Text Mining untuk Analisis Sentimen Terhadap Program Makan Siang Gratis di Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM)", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2024 Publication	1%
4	Submitted to Universitas Budi Luhur Student Paper	1%
5	Submitted to Konsorsium Perguruan Tinggi Swasta Indonesia II Student Paper	1%
6	Submitted to Universitas Slamet Riyadi Student Paper	1%
7	ojs.ninetyjournal.com Internet Source	<1%
8	123dok.com Internet Source	<1%
	etd.umy.ac.id	