

**ANALISIS SENTIMEN MASYARAKAT INDONESIA
TERHADAP PROGRAM MAKAN BERGIZI GRATIS
MENGUNAKAN *MULTILINGUAL BERT* DAN TEKNIK
*LABELLING TRANSFORMER***

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Sistem Informasi



disusun oleh

RABIATUL ADAWIYAH

23.22.2529

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**ANALISIS SENTIMEN MASYARAKAT INDONESIA
TERHADAP PROGRAM MAKAN BERGIZI GRATIS
MENGUNAKAN *MULTILINGUAL BERT* DAN TEKNIK
*LABELLING TRANSFORMER***

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Sistem Informasi



disusun oleh

RABIATUL ADAWIYAH

23.22.2529

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN MASYARAKAT INDONESIA
TERHADAP PROGRAM MAKAN BERGIZI GRATIS
MENGUNAKAN *MULTILINGUAL BERT* DAN TEKNIK
*LABELLING TRANSFORMER***

yang disusun dan diajukan oleh

RABIATUL ADAWIYAH

23.22.2529

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 31 Juli 2025

Dosen Pembimbing,



Anggit Dwi Hartanto M.Kom

NIK. 190302163

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN MASYARAKAT INDONESIA
TERHADAP PROGRAM MAKAN BERGIZI GRATIS
MENGUNAKAN *MULTILINGUAL BERT* DAN TEKNIK
*LABELLING TRANSFORMER***

yang disusun dan diajukan oleh

Rabiatul Adawiyah

23.22.2529

Telah dipertahankan di depan Dewan Penguji
pada tanggal 31 Juli 2025

Susunan Dewan Penguji

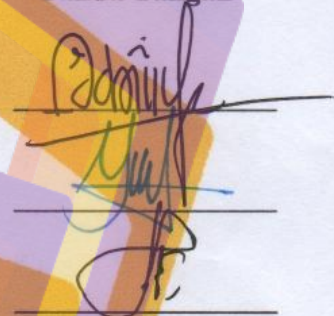
Nama Penguji

Tanda Tangan

Irma Rofni Wulandari, S.Pd., M.Eng.
NIK. 190302329

Yoga Pristyanto.S.Kom., M.Eng.
NIK. 190302412

Anggit Dwi Hartanto. S.Kom., M.Kom.
NIK. 190302163



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 31 Juli 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : RABIATUL ADAWIYAH
NIM : 23.22.2529

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Masyarakat Indonesia Terhadap Program Makan Bergizi Gratis Menggunakan Multilingual BERT dan Teknik Labelling Transformer

Dosen Pembimbing : Anggit Dwi Hartanto M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 31 Juli 2025

Yang Menyatakan,



Rabiatul Adawiyah

HALAMAN PERSEMBAHAN

Dengan segala kerendahan hati dan rasa syukur yang tiada terhingga atas kehadiran Allah SWT dan limpahan rahmat, kasih sayang, kekuatan, serta petunjuk-Nya, sehingga penulis dapat menyelesaikan skripsi ini sebagai salah satu syarat untuk meraih gelar Sarjana di Program Studi Sistem Informasi, Universitas Jogja.

Skripsi ini penulis persembahkan dengan tulus kepada:

1. Allah SWT, atas segala rahmat, hidayah, dan kasih sayang-Nya yang tiada henti dalam setiap langkah kehidupan penulis.
2. Kedua orang tua tercinta, Ibu dan Bapak penulis, yang selalu menjadi sumber kekuatan dan inspirasi dalam hidup. Terima kasih atas cinta yang tulus, doa yang tiada henti, serta pengorbanan tanpa batas yang telah dicurahkan hingga kini. Tanpa restu dan dukungan mereka, penulis tidak akan sampai pada titik ini. Mereka adalah alasan terbesar penulis bertahan dan terus berjuang.
3. Dosen pembimbing dan seluruh dosen Program Studi Sistem Informasi, Yang telah memberikan ilmu, arahan, serta bimbingan akademik dan moral selama masa studi penulis, khususnya dalam penyusunan skripsi ini. Ucapan terima kasih yang sebesar-besarnya penulis sampaikan atas segala kontribusi yang telah diberikan.
4. Seluruh sahabat dan rekan kerja, Yang telah menjadi bagian penting dalam kehidupan penulis, memberikan motivasi, dukungan, dan semangat selama proses perkuliahan hingga terselesaikannya skripsi ini.
5. Diri penulis sendiri, Sebagai bentuk apresiasi atas ketekunan, kerja keras, serta keteguhan dalam menghadapi berbagai tantangan selama proses penulisan skripsi ini.

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, karunia, dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul "Sentimen Analisis Masyarakat Indonesia Terhadap Makan Bergizi Gratis Menggunakan Multilingual BERT dan Pelabelan Transformer" sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Dalam proses penyusunan skripsi ini, penulis banyak memperoleh bantuan, bimbingan dan dukungan dari berbagai pihak baik secara langsung maupun tidak langsung. Oleh karena itu, dengan penuh hormat penulis mengucapkan terimakasih dan semoga Allah memberi balasan terbaik.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, baik dari segi isi maupun penyajiannya. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan demi perbaikan di masa mendatang.

Akhir kata, penulis berharap agar skripsi ini dapat memberikan manfaat bagi semua pihak yang berkepentingan, serta dapat menjadi referensi dan kontribusi nyata dalam pengembangan ilmu pengetahuan, khususnya di bidang Sistem Informasi.

Yogyakarta, 20 April 2025

Penulis

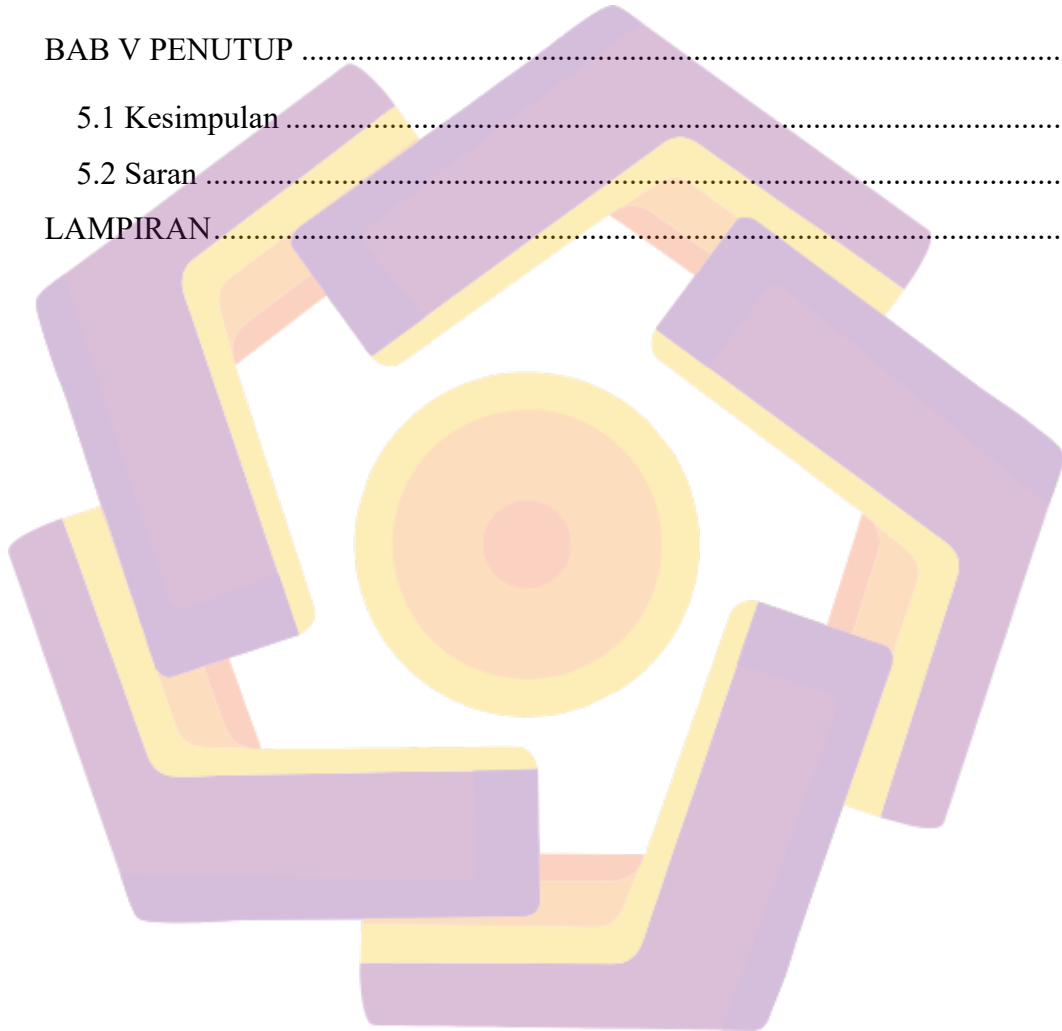
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR ISTILAH	xiii
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	12
2.2.1 <i>Machine Learning</i>	12
2.2.2 <i>Deep Learning</i>	12
2.2.3 <i>Natural Language Processing (NLP)</i>	12

2.2.4 Analisis Sentimen	13
2.2.5 <i>Text Mining</i>	13
2.2.6 Transformer.....	14
2.2.7 BERT	15
2.2.8 <i>Multilingual BERT</i>	16
2.2.9 <i>Preprocessing Data</i>	16
2.2.10 <i>Random Over Sampler</i>	17
2.2.12 K-Cross Fold Validation	17
2.2.13 <i>Confusion Matrix</i>	17
2.2.14 Instagram.....	19
2.2.15 Twitter/X	19
2.2.16 Program Makan Bergizi Gratis	20
BAB III METODE PENELITIAN	21
3.1 Objek Penelitian.....	21
3.2 Alur Penelitian	21
3.3 Akuisisi Data.....	22
3.3.1 Data Tweet X.....	22
3.3.2 Data Komentar Instagram	22
3.3.3 Data Normalisasi.....	23
3.4 Pre-processing Text	24
3.4.1 Case Folding	24
3.4.2 Cleaning	24
3.4.3 Normalisasi	24
3.4.4 <i>Labeling</i>	25
3.4.5 <i>Tokenizing</i>	25
3.4.6 <i>Stopword</i>	25

3.4.7 Stemming.....	25
3.5 Wordcloud.....	25
3.6 Random Over Sampler.....	26
3.7 Encoding.....	26
3.8 Splitting Data.....	26
3.9 Multilingual BERT (Bidirectional Encoder Representations from Transformers).....	27
3.10 K-Fold Cross Validation.....	27
3.11 Evaluasi.....	28
BAB IV HASIL DAN PEMBAHASAN.....	29
4.1 Akuisisi Data.....	29
4.2 Data Komentar Instagram.....	29
4.2.1 Data Tweet.....	30
4.2.2 Data Normalisasi.....	31
4.3 Pre-Processing.....	32
4.3.1 Case Folding.....	32
4.3.2 Cleaning.....	33
4.3.3 Normalisasi.....	35
4.3.4 Labelling.....	37
4.3.5 Tokenizing.....	39
4.3.6 Stopword.....	40
4.3.7 Stemming.....	41
4.4 WordCloud.....	41
4.5 Random Over Sampler.....	43
4.6 Encoder.....	43
4.7 Tokenizing.....	44
4.8 Splitting Data.....	47
4.9 Implementasi MBERT dan K-Fold Cross Validation.....	47

4.10 Evaluasi <i>Multilingual BERT</i>	48
4.10.1 Evaluasi menggunakan <i>k-fold</i> dan <i>Random Over Sampler</i>	48
4.10.2 Evaluasi menggunakan <i>Random Over Sampler</i> Tanpa <i>K-Fold</i>	52
4.10.3 Evaluasi menggunakan <i>K-Fold</i> Tanpa <i>Random Over Sampler</i>	54
4.10.4 Evaluasi tanpa menggunakan <i>K-fold</i> dan <i>Random Over Sampler</i>	56
BAB V PENUTUP	60
5.1 Kesimpulan	60
5.2 Saran	61
LAMPIRAN.....	66



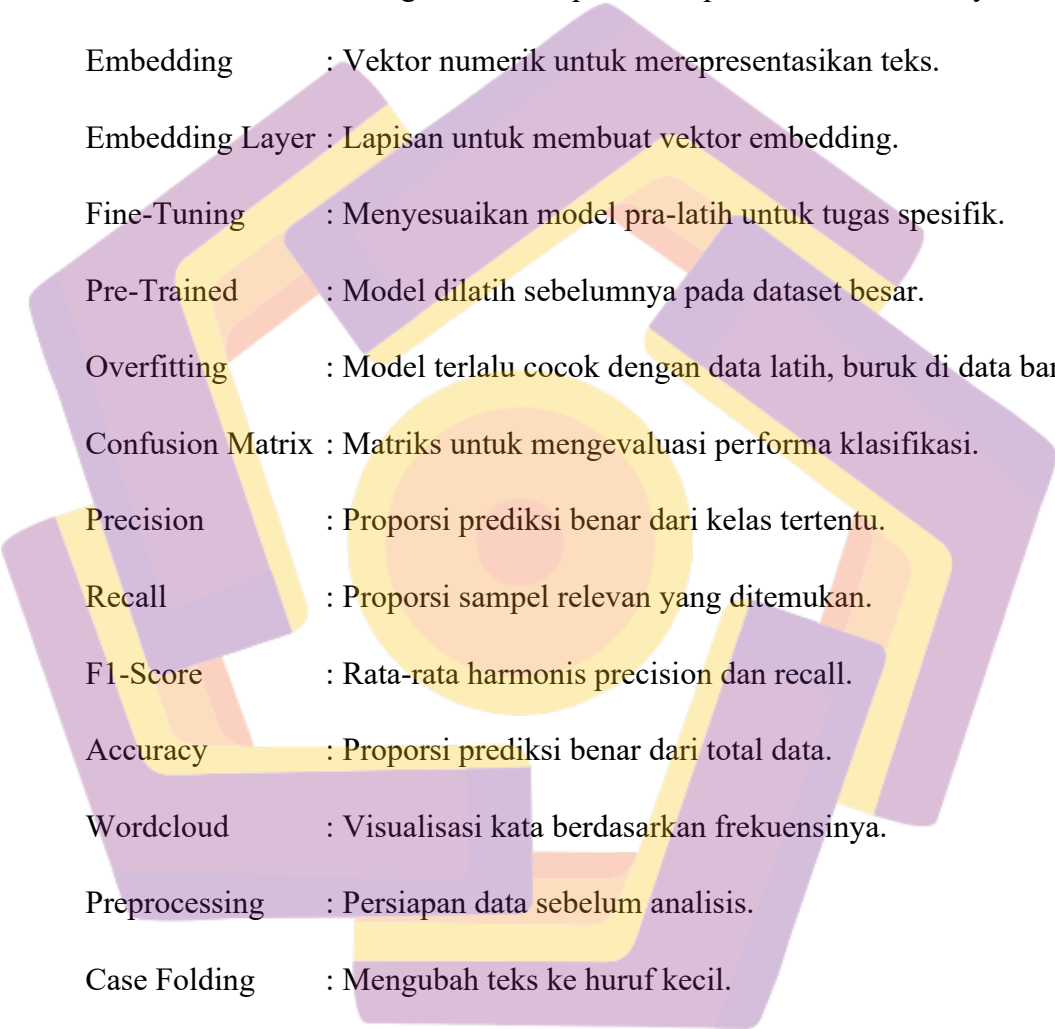
DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	9
Tabel 2. 2 Confusion Matrix	18
Tabel 3. 1 Sumber data Instagram	22
Tabel 4. 1 Dataset sebelum preprocessing	31
Tabel 4. 2 Dataset Normalisasi	32
Tabel 4. 3 Hasil Penerapan Case Folding	33
Tabel 4. 4 Hasil Penerapan Cleaning	35
Tabel 4. 5 Hasil tahapan normalisasi	36
Tabel 4. 6 <i>Tokenizing</i>	40
Tabel 4. 7 Proses <i>Tokenizing</i>	45
Tabel 4. 8 Hasil <i>Encoding, Padding, dan Attention Mask</i>	46
Tabel 4. 9 <i>Hyperparameter Multilingual BERT</i>	47
Tabel 4. 10 Tabel rata-rata <i>training loss</i> dengan <i>k-fold</i> dan <i>random over sampler</i>	48
Tabel 4. 11 Rata-rata <i>training loss</i> dengan <i>random over sampler</i> tanpa <i>k-fold</i>	52
Tabel 4. 12 Rata-rata <i>training loss</i> dengan <i>k-fold</i> tanpa <i>random over sampler</i>	54

DAFTAR GAMBAR

Gambar 4. 1 Informasi dataset normalisasi.....	36
Gambar 4. 2 Informasi data awal dan dataset setelah proses normalisasi	37
Gambar 4. 3 <i>Visualisasi</i> label <i>positive</i> dan <i>negative</i>	39
Gambar 4. 4 <i>Visualisasi wordcloud</i> sentimen <i>negative</i>	42
Gambar 4. 5 <i>Visualisasi wordcloud</i> sentimen <i>positive</i>	42
Gambar 4. 6 <i>Visualisasi balancing</i> dengan <i>random sampler</i> data <i>negative</i> dan <i>positive</i>	43
Gambar 4. 7 Hasil <i>encode</i>	44
Gambar 4. 8 <i>Confusion Matriks model</i> dengan <i>k-fold</i> dan <i>random over sampler</i> .	51
Gambar 4. 9 <i>Confusion Matriks</i> model dengan <i>k-fold</i> tanpa <i>random over sampler</i>	53
Gambar 4. 10 <i>Confusion Matriks</i> Model dengan <i>k-fold</i> tanpa <i>over sampler</i>	55
Gambar 4. 11 <i>Confusion matriks</i> model tanpa menggunakan <i>random over sampler</i> dan <i>k-fold</i>	58

DAFTAR ISTILAH



Self-Attention	: Memahami hubungan antar-kata dalam teks.
Encoder	: Mengolah input menjadi representasi tersembunyi.
Decoder	: Menghasilkan output dari representasi tersembunyi.
Embedding	: Vektor numerik untuk merepresentasikan teks.
Embedding Layer	: Lapisan untuk membuat vektor embedding.
Fine-Tuning	: Menyesuaikan model pra-latih untuk tugas spesifik.
Pre-Trained	: Model dilatih sebelumnya pada dataset besar.
Overfitting	: Model terlalu cocok dengan data latih, buruk di data baru.
Confusion Matrix	: Matriks untuk mengevaluasi performa klasifikasi.
Precision	: Proporsi prediksi benar dari kelas tertentu.
Recall	: Proporsi sampel relevan yang ditemukan.
F1-Score	: Rata-rata harmonis precision dan recall.
Accuracy	: Proporsi prediksi benar dari total data.
Wordcloud	: Visualisasi kata berdasarkan frekuensinya.
Preprocessing	: Persiapan data sebelum analisis.
Case Folding	: Mengubah teks ke huruf kecil.
Word Normalization	: Mengubah kata ke bentuk baku.
Tokenizing	: Memecah teks jadi kata/token.
Stopword Removal	: Menghapus kata-kata umum yang tidak penting.
Stemming	: Mengubah kata ke bentuk dasar.
AdamW	: Optimasi dengan regularisasi bobot.

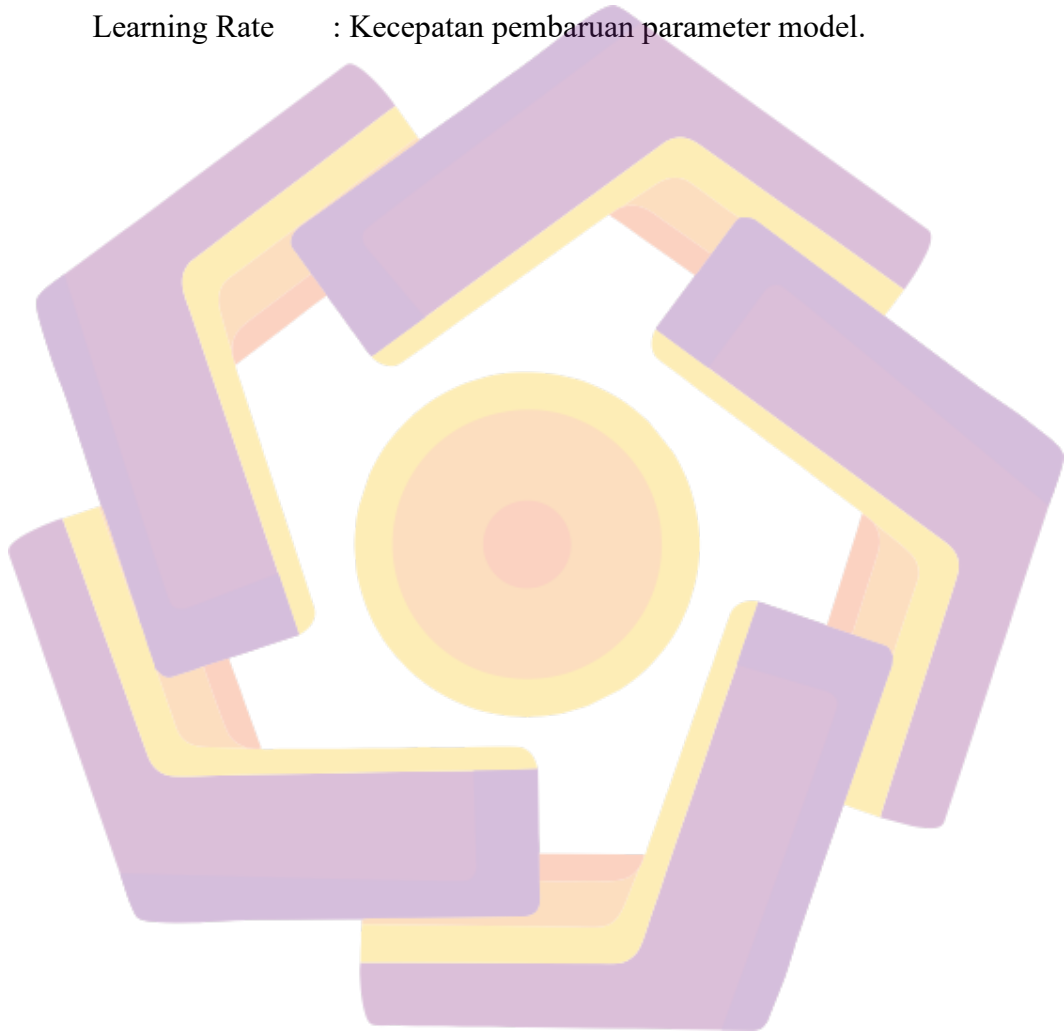
Hyperparameter : Parameter yang diatur sebelum pelatihan.

Epoch : Satu siklus pelatihan penuh.

Batch Size : Jumlah data per iterasi pelatihan.

Optimizer : Algoritma pembaruan bobot.

Learning Rate : Kecepatan pembaruan parameter model.



INTISARI

Program Makan Bergizi Gratis merupakan kebijakan pemerintah yang bertujuan meningkatkan kualitas gizi masyarakat, khususnya anak-anak sekolah. Kebijakan ini mendapat berbagai tanggapan dari masyarakat yang disampaikan melalui media sosial seperti *X* dan Instagram. Tanggapan yang beragam ini menimbulkan ketidakpastian dalam persepsi publik terhadap program makan Bergizi gratis, yang berpotensi memengaruhi keputusan pemangku kebijakan jika tidak dianalisis secara tepat. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program tersebut menggunakan pendekatan deep learning, yaitu model *Multilingual BERT*, serta teknik pelabelan otomatis berbasis arsitektur *transformer*. Sebanyak 19.128 data komentar dikumpulkan melalui teknik *scraping* dari media sosial *Instagram* dan *X*, kemudian dilakukan tahap *preprocessing*, pelabelan sentimen, dan penyeimbangan data menggunakan *Random over-sampling Technique*. Validasi model dilakukan menggunakan metode *K-Fold Cross Validation* untuk meningkatkan keandalan dan menghindari *overfitting*. Hasil analisis menunjukkan bahwa model mampu mengklasifikasikan sentimen secara efektif, dengan hasil akhir berupa 7.301 sentimen negatif dan 4.019 sentimen positif. Data yang sudah mendapatkan label uji pada ke empat skema yang dilakukan evaluasi Model *Multilingua BERT* yang diimplementasikan dengan *random over sampler* dan *k-fold* akurasi mencapai 99%, evaluasi dengan *k-fold tanpa over sampler* pada tahap ini menunjukkan akurasi sebesar 96,19% dan hasil evaluasi dengan *random over sampler* tanpa *k-fold* menunjukkan akurasi sebesar 90% sedangkan evaluasi model jika tanpa menggunakan *random over sampler* dan *k-fold* hanya mencapai akurasi 86%. kombinasi *Random over-sampling* dan *K-Fold* menghasilkan akurasi yang lebih unggul dan Penelitian ini memberikan gambaran representatif mengenai opini masyarakat dan diharapkan dapat menjadi landasan dalam evaluasi kebijakan, serta referensi untuk penelitian lebih lanjut dalam bidang analisis sentimen terhadap kebijakan publik menggunakan *multilingual BERT*.

Kata kunci: Sentimen Analisis, Makan Bergizi Gratis, *Multilingual BERT*, *Transformer*, NLP

ABSTRACT

Makan Bergizi Gratis is a government policy aimed at improving the nutritional quality of society, particularly school-aged children. This policy has received diverse responses from the public, which are expressed through social media platforms such as X and Instagram. These varying responses create uncertainty in public perception of the Makan Bergizi Gratis program, which may influence policymakers' decisions if not properly analyzed. This study aims to analyze public sentiment toward the program using a deep learning approach, specifically the Multilingual BERT model combined with an automated labeling technique based on the transformer architecture. A total of 19,128 comments were collected through web scraping from Instagram and X, followed by preprocessing, sentiment labeling, and data balancing using the Random Over-Sampling Technique. Model validation was performed using K-Fold Cross Validation to enhance reliability and prevent overfitting. The results show that the model effectively classifies sentiment, with a final distribution of 7,301 negative sentiments and 4,019 positive sentiments. The evaluation of the labeled data under four different scenarios demonstrates that the Multilingual BERT model, when implemented with Random Over-Sampling and K-Fold, achieved an accuracy of 99%. Evaluation with K-Fold without oversampling achieved an accuracy of 96.19%, while Random Over-Sampling without K-Fold achieved 90%, and evaluation without both techniques resulted in 86% accuracy. These findings indicate that the combination of Random Over-Sampling and K-Fold yields superior performance. This research provides a representative overview of public opinion and is expected to serve as a foundation for policy evaluation as well as a reference for further research on sentiment analysis of public policies using Multilingual BERT.

Keyword: Sentiment Analysis, Multilingual BERT, Transformer, Free Nutritious Meal Program, NLP