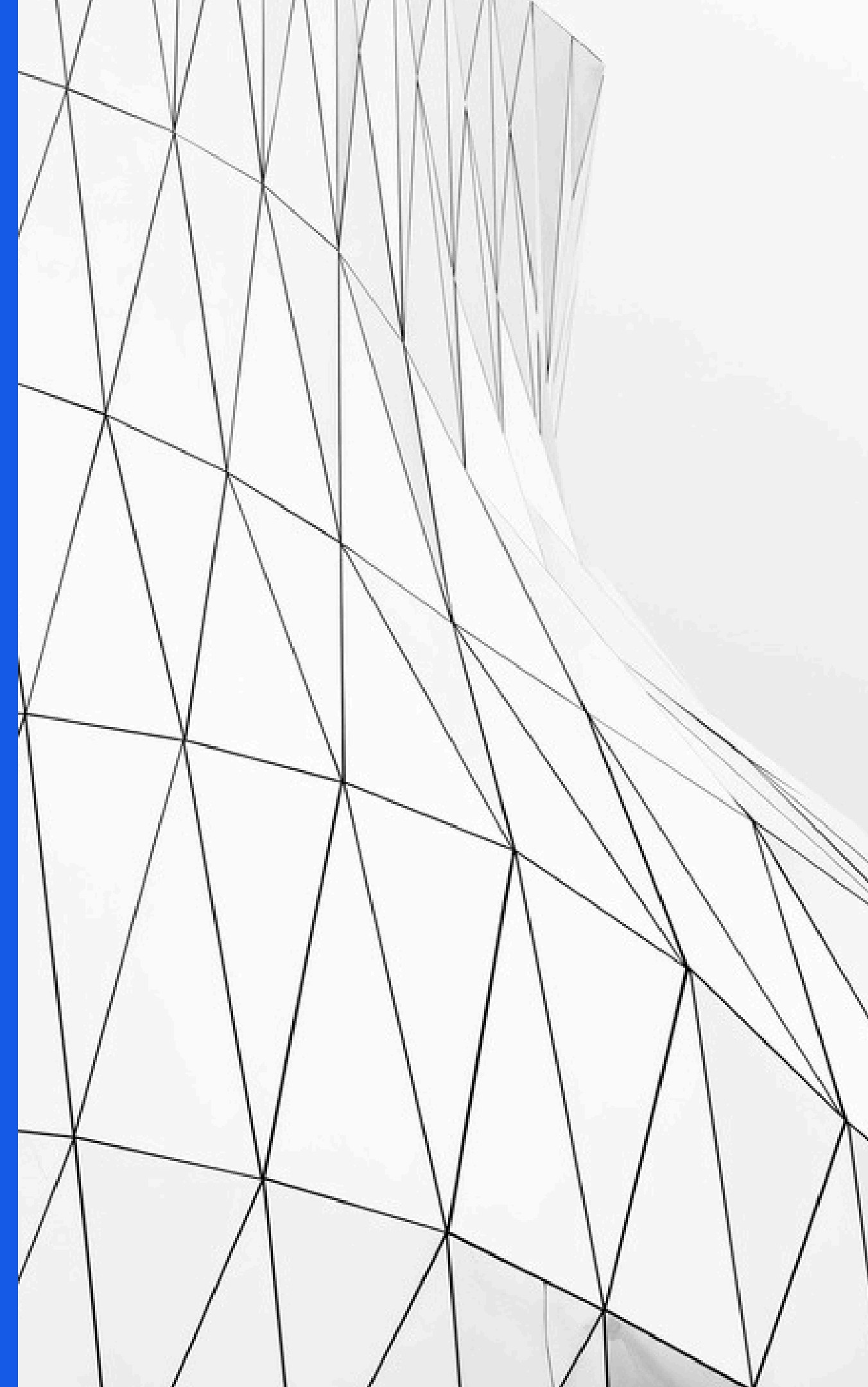


Analisis Sentimen Program MBG

Natural Language Processing • Twitter/X

oleh :

- Annisa Maisaroh (2304220003)
- Muhammad Zidane Yanfa Negara (2304220004)
- Maheswari Rahma Sarvidya (2304220019)
- Virgiawan Malik Rizky (2304220027)
- Moh. Farkhan Fauzan (2304220028)
- Ali Maulana (2304220031)
- Cindy Pramudita (2304220040)



Latar Belakang & Tujuan

Latar Belakang

- MBG jadi topik hangat di Twitter, memicu pro-kontra-netral.
- Twitter dimanfaatkan sebagai sumber opini publik real-time.
- Keterbatasan model stance detection Indo → digunakan pendekatan sentiment-based stance.

Tujuan

- Memetakan kecenderungan stance publik terhadap MBG.
- Menganalisis sentimen dengan Indo-RoBERTa.
- Mengidentifikasi isu dominan melalui visualisasi teks.



Dataset dan Preprocessing

Data dan Tahap Awal

Sumber: Tweet terkait Program MBG hasil crawling Twitter/X

Total data setelah cleaning & deduplikasi: 69 tweet

Data difilter berdasarkan duplikasi konten, bukan ID tweet.

Tahapan Preprocessing

1. Lowercasing
2. Hapus URL, mention, hashtag, simbol
3. Normalisasi spasi & karakter
4. Penghapusan stopwords Bahasa Indonesia

Output:
Kolom clean_text sebagai input utama untuk analisis NLZ

Model & Metodologi Analisis

Model & Pendekatan:

Indonesian RoBERTa Sentiment Classifier ([w11wo/indonesian-roberta-base-sentiment-classifier](https://huggingface.co/w11wo/indonesian-roberta-base-sentiment-classifier))

Alur Analisis:

1. Sentiment Analysis

Output model: positive, neutral, negative

2. Stance Mapping

Konversi sentimen ke stance:

positive → pro

negative → kontra

neutral → netral

3 Alasan Pemilihan Metode:

1. Model stance detection Bahasa Indonesia masih terbatas

2. Pendekatan sentiment-based stance → lebih stabil & praktis

3. Cocok untuk eksplorasi opini publik berskala besar

A bar chart titled 'Jumlah Tweet' (Number of Tweets) on the y-axis and 'Stance' on the x-axis. The y-axis ranges from 0 to 40 with increments of 5. There are three bars: a green bar for 'pro' with a value of 16 (23.19%), a grey bar for 'netral' with a value of 40 (57.97%), and a red bar for 'kontra' with a value of 13 (18.84%).

Stance	Jumlah Tweet	Percentage
pro	16	23.19%
netral	40	57.97%
kontra	13	18.84%



Hasil & Temuan Utama

Sentimen & Stance:

- **Netral paling dominan** → mayoritas tweet berupa informasi.
- **Positif**: manfaat gizi, ekonomi lokal, pemerataan pangan.
- **Negatif**: isu “tak perlu ahli gizi”, keracunan, gaji dapur, penyimpangan.
- **Metode sentiment-based stance inference** efektif digunakan sebagai pendekatan awal untuk studi stance detection pada data Bahasa Indonesia.

Word Cloud Insight:

Kata dominan: mbg, gizi, ahli gizi, sppg, anak, ekonomi, keracunan.

→ **Fokus publik** pada isu gizi, kontroversi kebijakan, dan teknis pelaksanaan.





Zero-Shot Classification (Model: BART)

Apa itu Zero-Shot Classification?

Zero-Shot Classification adalah metode klasifikasi teks tanpa pelatihan (training) khusus untuk label tersebut.

Cara Kerja:

Model yang sudah dilatih pada Natural Language Inference (NLI) akan:

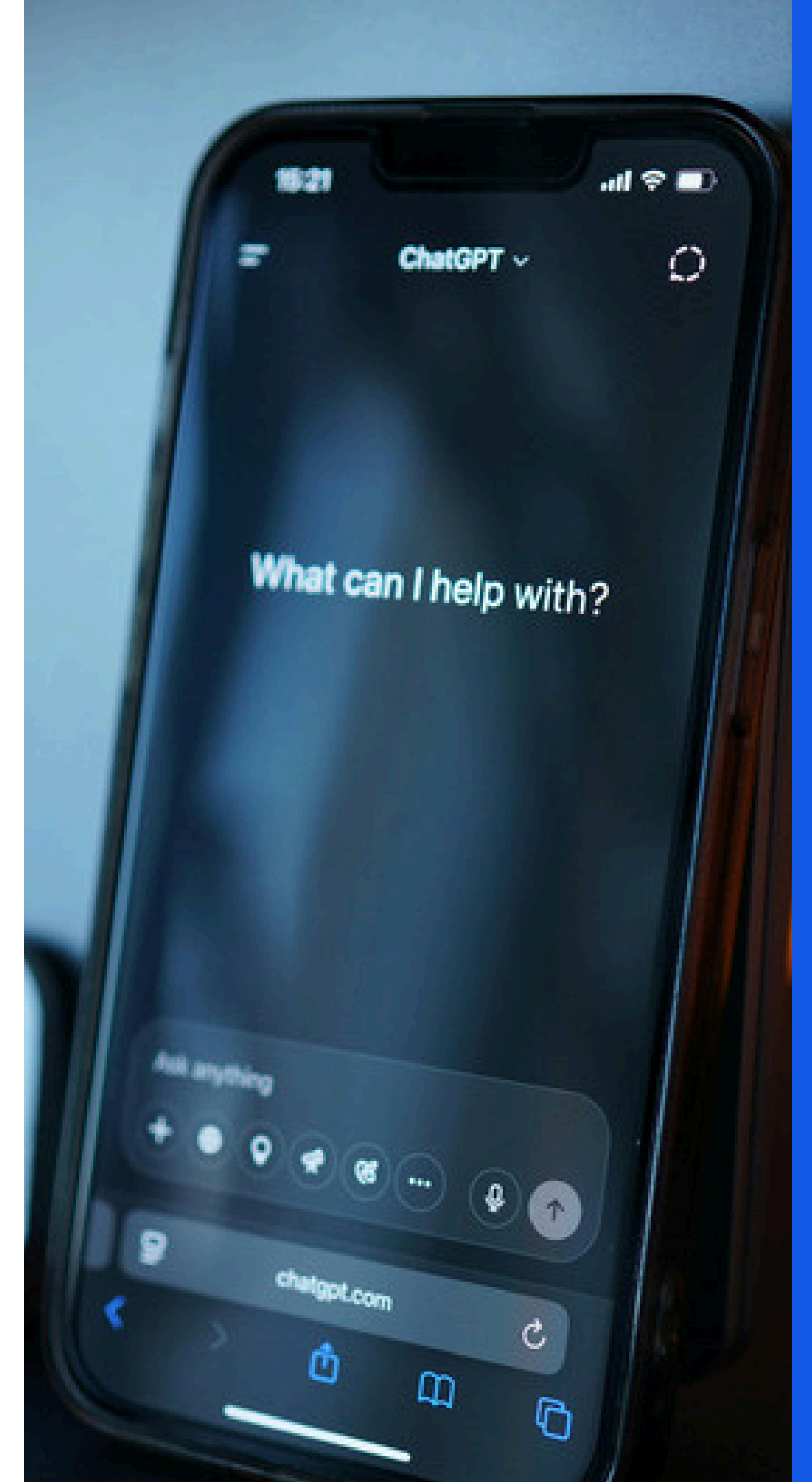
1. Mengambil teks sebagai premise
2. Mengubah label menjadi kalimat hypothesis
3. Menilai apakah teks tersebut sesuai dengan label itu (entailment), netral, atau bertentangan.

Apa itu BART?

BART atau *Bidirectional and Auto-Regressive Transformers* adalah model NLP yang menggabungkan kemampuan dari BERT (encoder, memahami konteks dua arah) dan GPT (decoder, menghasilkan teks)

Struktur:

- Encoder: memahami makna
- Decoder: membuat prediksi



Alur Pengerjaan

Link COLAB: [Klik di sini](#)

Data dan Tahap Awal

1. Library: Mengimpor library standar
2. Akses Google Drive: Melakukan mounting Google Drive untuk mengakses file data.
3. Impor Data Mentah: Membaca data tweet mentah dari file Raw_Crawling_MBG_X.txt. Data ini merupakan serangkaian raw text dengan format yang terstruktur (memiliki baris-baris berlabel seperti Date, Tweet ID, Text, Language, dan counts).

Tahapan Preprocessing

1. Lowercasing
2. Hapus URL, mention, hashtag, simbol
3. Normalisasi spasi & karakter
4. Penghapusan stopwords Bahasa Indonesia

```
# daftar stopwords indonesia
stopwords_id = set([
    "dan", "di", "yang", "ke", "dari", "untuk", "pada", "dengan", "itu", "ini",
    "sebagai", "karena", "atau", "juga", "ada", "agar", "dalam", "nya", "buat",
    "kita", "kamu", "saya", "aku", "dia", "mereka", "kami", "pun", "ya", "nah",
    "si", "para", "tersebut", "jadi", "bagi"
])
```

before: 99

after : 69

removed: 30

Melakukan NLP dengan Metode Zero-Shot Classification (Model: BART)

```
import pandas as pd
from transformers import pipeline

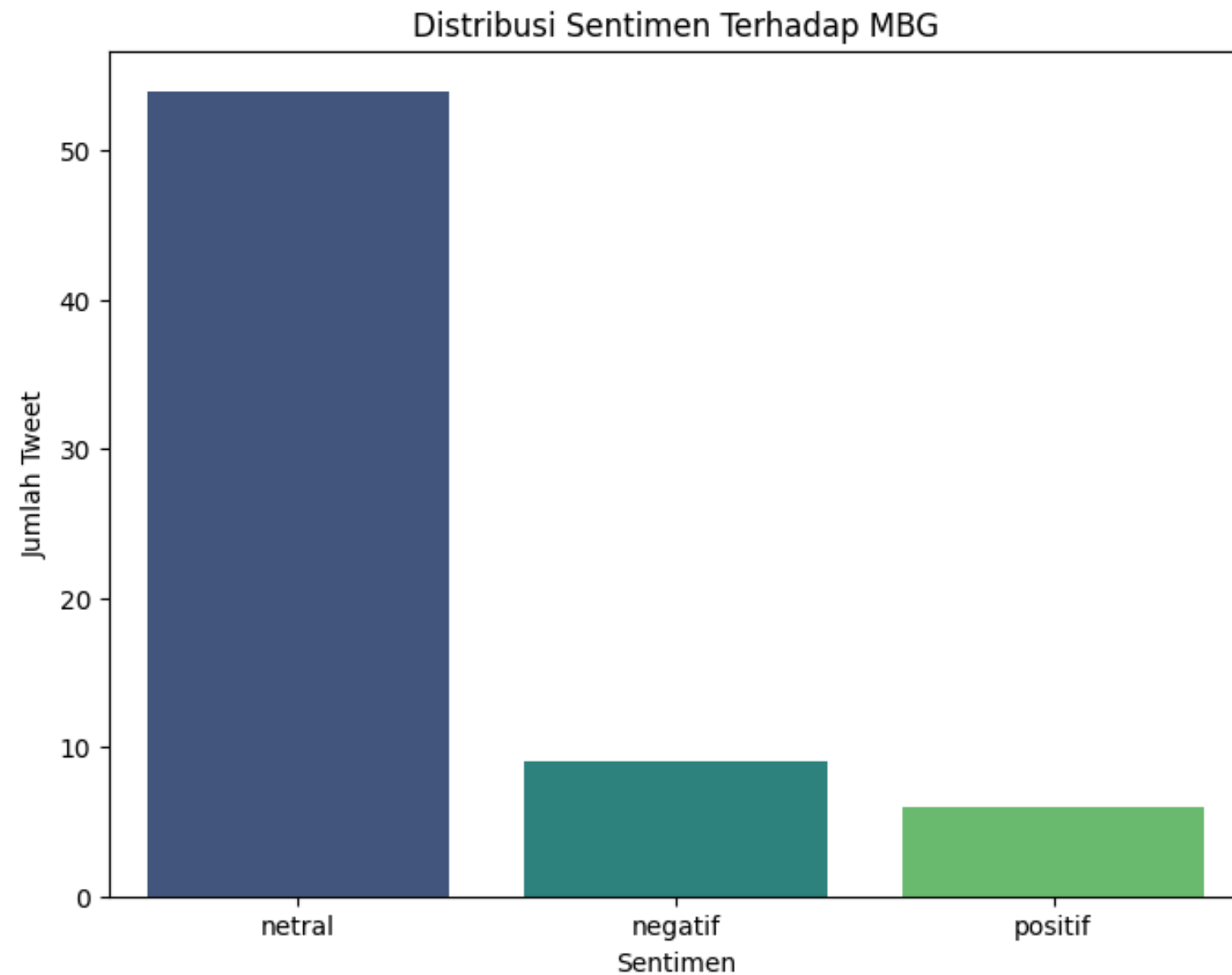
# 1. Tentukan Model dan Pipeline
# Zero-shot classification memungkinkan model (seringkali model NLI seperti BART atau RoBERTa)
# untuk mengklasifikasikan teks berdasarkan label yang diberikan tanpa training spesifik.
# Saya akan menggunakan model umum untuk zero-shot classification.
classifier = pipeline(
    "zero-shot-classification",
    model="facebook/bart-large-mnli"
)

# 2. Tentukan Label Sentimen
# Label yang akan digunakan untuk klasifikasi.
candidate_labels = ["positif", "negatif", "netral"]

# 3. Definisikan Fungsi Klasifikasi
def analyze_sentiment_zero_shot(text, classifier, labels):
    if not text:
        return {"sentiment": "netral", "score": 0.0} # Handle empty text
    try:
        result = classifier(text, labels)
        # Ambil label dengan skor tertinggi
        sentiment = result['labels'][0]
        score = result['scores'][0]
        return {"sentiment": sentiment, "score": score}
    except Exception as e:
        print(f"Error classifying text: {text}. Error: {e}")
        return {"sentiment": "error", "score": 0.0}
```

Tabel Hasil Analisis Hasil Analisis

	Tweet ID	clean_text	sentiment	score
0	1990828187006201982	kepala bgn ungkap penggunaan nitrin tinggi pen...	netral	0.499916
1	1990808795518546193	program mbg dorong pertumbuhan umkm tempe perm...	netral	0.553383
2	1990807038474006730	harta kekayaan cucun ahmad wakil ketua dpr per...	netral	0.433071
3	1990791277802394082	polemik mbg makin panas ahli gizi menyampaikan...	netral	0.552708
4	1990790947966447998	presiden brazil memerlukan 11 tahun mencapai 4...	netral	0.577597



Interpretasi dan Kesimpulan

1. Dominasi Sentimen Netral

Sebagian besar diskusi di Twitter didominasi oleh sentimen Netral yaitu sebanyak 54 tweet atau sekitar 78.26%. Hal ini mungkin disebabkan oleh banyaknya tweet berita yang bersifat netral.

2. Isu Sentimen Negatif

Meskipun netral mendominasi, terdapat sejumlah sentimen negatif. Berdasarkan contoh yang ditampilkan, sentimen negatif ini banyak berpusat pada polemik seputar program MBG.

3. Topik Utama (Word Cloud)

Kata-kata yang paling sering dibicarakan adalah:

- MBG dan Makan Bergizi Gratis.
- Pihak-pihak terkait, seperti Ahli Gizi, DPR, dan BGN (Badan Gizi Nasional).
- Isu-isu seperti Program, Stunting, Ekonomi, Kesehatan, dan Tumbuh.



RoBERTa-base

([cardiffnlp/twitter-roberta-base-sentiment-latest](#))



Apa itu RoBERTa-base?

RoBERTa-base adalah model bahasa buatan Facebook AI yang merupakan versi penyempurnaan dari BERT. Model ini dilatih dengan data jauh lebih banyak, tanpa tugas Next Sentence Prediction, dan dengan pengaturan training yang lebih optimal sehingga performanya biasanya lebih baik daripada BERT standar. Ukuran “base” berarti model ini memakai arsitektur 12 layer dengan sekitar 125 juta parameter—cukup ringan tapi tetap kuat untuk tugas NLP seperti klasifikasi teks, sentiment analysis, dan question answering.

Bagaimana Cara Kerjanya?

RoBERTa-base bekerja dengan membaca teks lalu memecahnya menjadi potongan kata kecil, mengubahnya menjadi vektor angka, dan memprosesnya melalui rangkaian Transformer encoder yang memakai mekanisme self-attention supaya setiap kata bisa “melihat” konteks kata lain dalam kalimat. Saat pre-training, sebagian kata sengaja disembunyikan dan RoBERTa belajar menebaknya, sehingga model memahami pola bahasa. Setelah itu, model bisa dilatih ulang untuk tugas tertentu seperti sentiment analysis atau klasifikasi teks dengan menambahkan layer sederhana di atasnya. Singkatnya, RoBERTa-base memahami makna kalimat dengan mempelajari hubungan antar kata dan konteksnya.

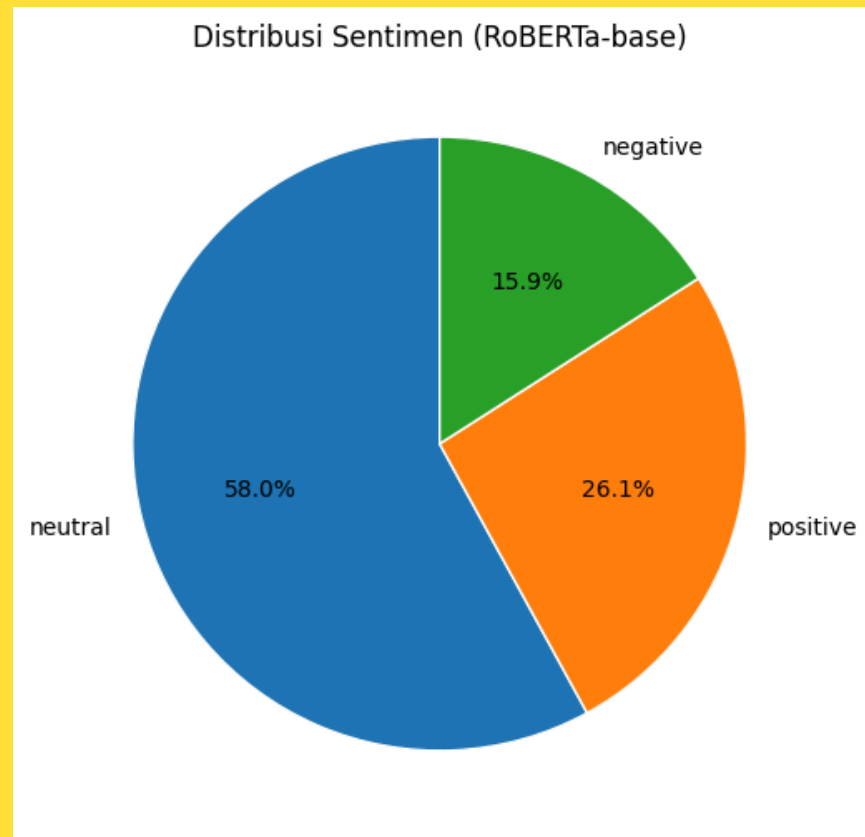
Keunggulan

- Dilatih dengan data lebih banyak sehingga akurasi umumnya lebih tinggi dari BERT.
- Tidak memakai Next Sentence Prediction, sehingga training lebih efisien dan hasilnya lebih stabil.
- Performa kuat di banyak tugas NLP seperti klasifikasi, dan sentiment analysis.
- Ukuran “base” masih cukup ringan sehingga fine-tuning relatif cepat dan tidak terlalu boros sumber daya.
- Representasi konteksnya lebih kaya berkat optimasi self-attention dan masking yang lebih agresif.

Kelemahan

- Tetap membutuhkan GPU atau hardware cukup kuat untuk training dan inferensi cepat.
- Versi aslinya hanya dilatih pada bahasa Inggris, sehingga kinerjanya menurun di bahasa lain tanpa fine-tuning khusus.
- Model cukup besar (125M parameter) sehingga kurang cocok untuk aplikasi edge atau perangkat rendah daya.
- Seperti model Transformer lainnya, tidak benar-benar memahami konteks dunia nyata—hanya belajar pola statistik dari data.

Hasil Analisis dan Visualisasi



0.7354596918043883

Interpretasi dan Kesimpulan

1. Distribusi Sentimen

Dari pie chart didapat:

- Neutral (58%) mendominasi. Ini berarti sebagian besar teks bersifat informatif, deskriptif, atau tidak menunjukkan emosi tertentu. Biasanya terjadi pada berita atau pernyataan resmi.
- Positive (26.1%) lumayan besar, menunjukkan cukup banyak narasi yang mendukung, mengapresiasi, atau memberikan respon positif, kemungkinan terkait program atau kebijakan.
- Negative (15.9%) relatif kecil, tapi tetap signifikan. Ini menunjukkan ada kritik, kekhawatiran, atau nada negatif dalam sebagian pembahasan.

Maknanya:

Wacana terkait topik #MBG, analisis cenderung moderat dan netral, dengan campuran dukungan cukup kuat dan kritik yang masih ada tapi tidak dominan.

Interpretasi dan Kesimpulan

2. Wordcloud

Interpretasinya:

- Pembahasan banyak terpusat pada program gizi anak, makanan, dapur umum, bantuan rakyat, dan kebijakan publik.
- Kata “dpr”, “ketua”, “wakil” mengindikasikan wacana politik atau komentar pejabat.
- Ada nada sosial & kesehatan: asupan, sehat, kualitas, tumbuh.
- Kata “maaf”, “tidak”, “bukan”, “arogan” menunjukkan ada konteks klarifikasi, kritik, atau respon negatif dalam sebagian berita.

Maknanya:

Diskusi publik (atau berita) tentang topik ini cenderung memuat campuran isu sosial—terutama gizi anak, kesejahteraan, pernyataan pejabat, dan tanggapan masyarakat.



Thank You