Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru Loktabat – Banjarbaru,
e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0893

Analisis Sentimen dan Distorsi Kognitif Konflik Anggaran Pendidikan dengan Makan Bergizi Gratis

DOI: http://dx.doi.org/10.35889/jutisi.v15i4.3924

Azzrial Arfiansyah[1*], Cinta Auliya Kusuma Ananda[2], Quyun Isnawan[3]
Sistem Informasi, Universitas Pembangunan Nasional "Veteran" Jakarta, Jakarta, Indonesia
**e-mail Corresponding Author*: azzriala@gmail.com

Abstract

The budget allocation conflict of the Free Nutritious Meal (MBG) program amounting to Rp223.5 trillion within the 2026 education budget has triggered opinion polarization and collective thinking bias on social media X. This study aims to analyze public sentiment distribution, identify dominant cognitive distortions, and test the statistical correlation between them. The methodology employs a two-layer parallel NLP pipeline utilizing the IndoBERTweet model for sentiment classification and the IndoBERT-base model for multilabel cognitive distortion detection across 2,000 tweets. Evaluation results show that the sentiment model achieves 83.25% accuracy, while the cognitive distortion detection reaches a subset accuracy of 83.75%, with Mental Filter being the most dominant type (8.2%). Fisher's Exact test proves a significant association (p < 0.05) between negative sentiment and the occurrence of Mental Filter and Catastrophizing. In conclusion, this hybrid approach effectively maps public collective thinking biases toward public policy dynamics.
Keywords: *Sentiment Analysis; Cognitive Distortion; Free Nutritious Meal; IndoBERTweet; Fleiss' Kappa*

Abstrak

Konflik alokasi anggaran program Makan Bergizi Gratis (MBG) sebesar Rp223,5 triliun dalam anggaran pendidikan APBN 2026 memicu polarisasi opini dan bias berpikir kolektif di media sosial X. Penelitian ini bertujuan menganalisis distribusi sentimen publik, mengidentifikasi distorsi kognitif dominan, serta menguji korelasi statistik di antara keduanya. Metodologi yang digunakan adalah *pipeline* NLP dua lapis paralel berbasis model IndoBERTweet untuk klasifikasi sentimen dan IndoBERT-base untuk deteksi distorsi kognitif multi-label terhadap 2.000 *tweet*. Hasil evaluasi menunjukkan model sentimen mencapai akurasi 83,25%, sedangkan deteksi distorsi kognitif meraih subset accuracy 83,75% dengan *Mental Filter* sebagai tipe paling dominan (8,2%). Uji *Fisher's Exact* membuktikan adanya hubungan signifikan (p < 0,05) antara sentimen negatif dengan kemunculan distorsi kognitif *Mental Filter* dan *Catastrophizing*. Simpulannya, pendekatan *hybrid* ini efektif dalam memetakan bias berpikir kolektif masyarakat terhadap dinamika kebijakan publik.
Kata Kunci: *Analisis Sentimen; Distorsi Kognitif; Makan Bergizi Gratis; IndoBERTweet; Fleiss' Kappa*

1. Pendahuluan

Pada Anggaran Pendapatan dan Belanja Negara (APBN) 2026 yang disahkan melalui Undang-Undang Nomor 17 Tahun 2025 [1], pemerintah mengalokasikan anggaran pendidikan sebesar Rp769,1 triliun guna memenuhi mandat konstitusional *mandatory spending* sebesar 20%. Berdasarkan Peraturan Presiden Nomor 118 Tahun 2025 tentang Rincian APBN TA 2026 [2], pendanaan operasional program Makan Bergizi Gratis (MBG) sebesar Rp223,5 triliun, atau sekitar 29% dari total anggaran pendidikan, dimasukkan ke dalam pos anggaran fungsi pendidikan. Dasar hukum dari inklusi anggaran ini adalah Pasal 22 ayat (3) *UU* 17/2025 beserta penjelasannya, yang menegaskan bahwa pendanaan operasional penyelenggaraan pendidikan mencakup program makan bergizi pada lembaga yang berkaitan dengan penyelenggaraan pendidikan. Secara operasional, penyelenggaraan program MBG ini dikelola oleh Badan Gizi

Nasional (BGN) berdasarkan Peraturan Presiden Nomor 115 Tahun 2025 tentang Tata Kelola Penyelenggaraan Program MBG [3].

Pengalihan dana dalam skala besar tersebut memicu persaingan anggaran *internal* (*intra-budgetary competition*) yang dikhawatirkan dapat mereduksi alokasi riil untuk operasional sekolah, dana Bantuan Operasional Sekolah (BOS), sarana prasarana, serta kesejahteraan guru. Benturan alokasi ini kemudian melahirkan tiga permohonan *Judicial Review* di Mahkamah Konstitusi oleh Koalisi Selamatkan Pendidikan Indonesia (KOSPI) yang diwakili oleh YLBHI/LBH Jakarta, yaitu melalui perkara Nomor 40, 52, dan 55/PUU-XXIV/2026 [4], yang menggugat konstitusionalitas Pasal 22 ayat (3) UU 17/2025. Polemik hukum dan kebijakan ini memicu gelombang opini publik di media sosial X (Twitter), yang telah menjadi medium representatif untuk mengukur sentimen publik terhadap kebijakan pemerintah di Indonesia [5], dan sering kali mencerminkan adanya distorsi kognitif (*cognitive distortion*), yaitu pola berpikir bias dan irasional yang dikaji dalam kerangka teori *Cognitive Behavioral Therapy* (*CBT*) [6], [7].

Beberapa riset terbaru telah berupaya menganalisis sentimen publik terhadap program MBG di media sosial Indonesia menggunakan model IndoBERT [8], [9], [10]. Namun, seluruh kajian tersebut hanya berfokus pada klasifikasi polaritas sentimen semata tanpa mengeksplorasi bias berpikir atau distorsi kognitif yang melandasi opini publik tersebut. Di sisi lain, penelitian mengenai deteksi distorsi kognitif secara komputasional baru berkembang pesat dalam beberapa tahun terakhir [11], [12], [13], [14], termasuk pembuatan *dataset* distorsi kognitif pertama berbahasa Indonesia oleh Sastra et al. [15], yang melengkapi ketersediaan sumber daya pemrosesan bahasa alami di Indonesia [16], [17]. Meskipun demikian, seluruh riset deteksi distorsi kognitif yang ada saat ini masih berada dalam domain klinis atau kesehatan mental, dan belum ada yang mengaplikasikannya pada domain kebijakan publik maupun fiskal. Hal ini menunjukkan adanya kesenjangan riset (*research gap*), di mana belum terdapat penelitian yang mengombinasikan analisis sentimen dengan deteksi distorsi kognitif secara paralel pada wacana kebijakan publik di Indonesia.

Untuk menjawab kesenjangan tersebut, penelitian ini mengusulkan pendekatan *pipeline Natural Language Processing* (NLP) dua lapis paralel (*two-layer parallel pipeline*). Lapis pertama untuk klasifikasi sentimen menggunakan IndoBERTweet [18], sedangkan lapis kedua untuk deteksi multi-label distorsi kognitif menggunakan IndoBERT-base. Penggunaan dua model terpisah (alih-alih *multi-task*) dipilih guna menghindari risiko *negative transfer* [19] mengingat sentimen (afektif) dan distorsi kognitif (penalaran) memiliki dasar linguistik yang berbeda. Pendekatan *task-specific fine-tuning* ini juga mencegah model didominasi oleh sinyal mayoritas akibat ketimpangan data kelas distorsi kognitif (hanya 12,5%), sehingga tiap model dapat mengoptimalkan parameter representasinya secara independen [20], [21]. Dari sisi teoretis, kelayakan operasionalisasi konstruk distorsi kognitif, yang secara tradisional dikaji dalam ranah psikologi klinis, ke dalam bentuk klasifikasi tekstual berbasis *transformer* telah dibuktikan pada berbagai studi terdahulu [11], [12], [13], [14], [15], yang menunjukkan bahwa pola linguistik distorsi kognitif dapat dikenali oleh model bahasa tanpa bergantung pada konteks klinis tertentu, sehingga terbuka peluang penerapannya pada domain wacana publik non-klinis seperti kebijakan fiskal. Dengan demikian, kebaruan (*novelty*) riset ini mencakup: (1) penerapan komputasional distorsi kognitif pada domain non-klinis (kebijakan fiskal); (2) penggunaan arsitektur BERT paralel independen pada satu dataset; dan (3) relevansi empiris isu mutakhir konflik APBN 2026.

Berdasarkan pemaparan latar belakang dan rumusan masalah tersebut, tujuan dari penelitian ini adalah: (1) menganalisis distribusi sentimen publik terhadap konflik anggaran pendidikan dengan program MBG; (2) mengidentifikasi tipe distorsi kognitif yang paling dominan muncul dalam narasi publik; dan (3) menguji korelasi statistik antara polaritas sentimen dengan kecenderungan timbulnya tipe distorsi kognitif tertentu.

2. Metodologi

Penelitian ini menggunakan metodologi terstruktur yang mencakup pengumpulan data wacana media sosial X (*Twitter*), pra-pemrosesan teks tanpa *stemming* untuk menjaga integritas semantik, anotasi manual oleh tiga *rater* independen, pengujian reliabilitas antar-*rater* menggunakan *Fleiss' Kappa*, pelatihan model *deep learning* berbasis *transformer* paralel, serta pengujian korelasi statistik menggunakan *Chi-Square* dan *Fisher's Exact*.

2.1 *Dataset* Penelitian

Dataset penelitian ini bersumber dari opini publik di platform X mengenai konflik anggaran pendidikan dan Program Makan Bergizi Gratis (MBG). Platform ini dipilih karena telah menjadi sumber data yang lazim digunakan pada penelitian sejenis mengenai wacana kebijakan publik Program MBG di Indonesia [8], [9], [22]. Proses *crawling* dilakukan menggunakan *framework* otomatisasi browser Playwright selama periode 1 Januari hingga 26 Juni 2026, bertepatan dengan memuncaknya polemik dan pengajuan *judicial review* terhadap APBN 2026 [1]. Untuk menjaring data yang komprehensif, ekstraksi dilakukan menggunakan 16 kata kunci spesifik yang mencakup isu MBG, anggaran pendidikan, Badan Gizi Nasional, guru honorer, dan hukum APBN. Seluruh *tweet* yang terkumpul kemudian melalui tahap penyaringan awal, meliputi penghapusan data duplikat, *tweet* kosong, dan konten tidak relevan, sebelum dilanjutkan ke tahap *text preprocessing*, anotasi sentimen dan distorsi kognitif, serta pemodelan klasifikasi.

2.2 Pra-Pemrosesan Data

Seluruh *tweet* yang diperoleh dari hasil *crawling* akan memasuki tahap *preprocessing* untuk meningkatkan kualitas data sebelum digunakan pada proses klasifikasi. Tahapan ini bertujuan menghilangkan elemen-elemen yang tidak memiliki makna semantik serta menyeragamkan format teks sehingga dapat diproses secara optimal oleh model *machine learning* [8], [10]. Proses ini dilakukan secara bertahap menggunakan bahasa pemrograman *Python* dengan memanfaatkan beberapa *library*, seperti *pandas, re, emoji,* dan Sastrawi.

Tahap pertama adalah *case folding*, yaitu mengubah seluruh karakter menjadi huruf kecil (*lowercase*) agar tidak terjadi perbedaan representasi antara kata yang memiliki makna sama. Selanjutnya dilakukan *cleaning* dengan menghapus *URL*, *username* (*@mention*), *hashtag* (#), angka, tanda baca berlebih, dan karakter khusus. *Emoji* dikonversi ke dalam deskripsi tekstual bahasa Indonesia untuk mempertahankan sinyal sentimennya. Setelah pembersihan, dilakukan *tokenization* untuk memisahkan teks menjadi *token*. *Token* yang dihasilkan kemudian melewati tahap *stopword removal* untuk menghilangkan kata-kata umum dengan informasi rendah, serta normalisasi kata tidak baku dilakukan dengan mengacu pada pemetaan bahasa gaul (*slang*) dan singkatan informal dalam analisis sentimen berbahasa Indonesia [23]. Proses *stemming* sengaja tidak dilakukan dalam penelitian ini. Karena model *BERT* menggunakan tokenisasi *WordPiece* yang secara inheren mampu menangani variasi morfologis kata, membiarkan imbuhan tetap utuh terbukti mempertahankan nuansa emosi dan struktur semantik yang sangat krusial untuk analisis sentimen dan pendeteksian distorsi kognitif [8], [15].

2.3 Anotasi Data dan Uji Reliabilitas

Dataset diberi label manual untuk dua tugas klasifikasi: analisis sentimen (positif, netral, negatif) dan deteksi *cognitive distortion* (*catastrophizing, overgeneralization, all-or-nothing thinking, mental filter,* dan *none*). Proses pelabelan mengacu pada pedoman standar [15], [12], serta mengacu pada skala distorsi kognitif yang tervalidasi [24].

Untuk menjaga objektivitas, pelabelan dilakukan secara independen oleh tiga anotator. Uji reliabilitas *inter-rater* dihitung menggunakan *Fleiss' Kappa* dengan persamaan berikut:

$$\kappa = \frac{P - P_e}{1 - P_e} \tag{1}$$

$$P = \frac{1}{N\,n\,(n-1)} \left[\sum_{i=1}^{N} \sum_{j=1}^{k} n_{ij}^2 - N\,n \right] \tag{2}$$

dimana κ menyatakan koefisien *Fleiss' Kappa*, P adalah proporsi kesepakatan aktual antar anotator, P_e adalah proporsi kesepakatan harapan acak, N adalah total data *tweet* (2.000), n adalah jumlah *rater* (3), k adalah jumlah kategori label, dan n_{ij} adalah jumlah *rater* yang memberikan label kategori j pada *tweet* ke-i.

Koefisien yang diperoleh adalah 0,9895 untuk label sentimen dan rentang 0,8675–0,9658 untuk kategori distorsi kognitif. Berdasarkan standar Landis & Koch, seluruh nilai tersebut termasuk dalam kategori *Almost Perfect Agreement* ($\kappa \geq 0{,}81$), yang secara mutlak memenuhi standar penilaian kesepakatan *inter-rater* [25], [26], [27], sekaligus memvalidasi objektivitas dan keandalan 2.000 data konsensus yang digunakan.

2.4 Model Klasifikasi

Penelitian ini mengusulkan arsitektur *Two-Layer Parallel Pipeline* menggunakan model *deep learning* berbasis *Transformer* seperti *BERT* yang telah terbukti andal dalam berbagai tugas pemrosesan bahasa alami [28], [29], [30]. Pada lapis 1 (Sentimen), digunakan model IndoBERTweet (*indolem/indobertweet-base-uncased*) yang dilatih menggunakan 5-*Fold Cross-Validation*, di mana prediksi akhir dihasilkan melalui model *ensemble* berbasis *rata-rata* probabilitas (*probability mean-pooling*). Pada lapis 2 (*Cognitive Distortion*), digunakan model IndoBERT-*base* (*indobenchmark/indobert-base-p1*) yang dilatih sebagai model tunggal (*Single Model*) untuk menghindari efek *signal drowning* dari perataan *ensemble* pada kelas minoritas ekstrem yang sangat langka. Lapis 2 dilatih dengan *Weighted Binary Cross Entropy* (*BCE*) menggunakan *parameter pos_weight* untuk menangani ketimpangan kelas yang sangat tinggi [20], [21].

$$L_c = -\frac{1}{N}\sum_{i=1}^{N} [\, w_c y_{ic}\, log(\sigma(x_{ic})) + (1 - y_{ic})\, log(1 - \sigma(x_{ic}))\,] \quad (3)$$

Di mana L_c menyatakan *loss* untuk kelas c, N menyatakan jumlah sampel data, y_{ic} menyatakan label biner aktual (0 atau 1) untuk kelas c pada data ke-i, x_{ic} menyatakan *output logit* dari model lapis 2, σ menyatakan fungsi aktivasi *sigmoid*, dan w_c menyatakan *parameter pos_weight* untuk bobot kelas positif.

Sebelum pelatihan dilakukan, *dataset* 2.000 *tweet* dibagi secara stratifikasi dengan proporsi 80:20 menjadi data latih+val (1.600 *tweet*) dan data uji *hold-out* (400 *tweet*). Teks diubah menjadi representasi numerik menggunakan *tokenizer* masing-masing model dengan batas panjang maksimal (*MAX_LEN*) 256. Kedua model dilatih menggunakan AdamW *optimizer* dengan *learning rate* 2e-5, *weight decay* 0.01, dan *warmup ratio* 0.1 selama maksimal 10 *epoch* dengan *batch size* 16. Teknik *early stopping* dengan *patience* 3 diterapkan berdasarkan *loss* validasi untuk menghindari *overfitting*.

2.5 Evaluasi Model

Evaluasi dilakukan menggunakan *confusion matrix* untuk melihat jumlah prediksi yang benar maupun salah pada setiap kelas. Berdasarkan hasil tersebut dihitung beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Keempat metrik tersebut digunakan karena mampu memberikan gambaran yang lebih komprehensif mengenai kemampuan model, terutama ketika distribusi data antar kelas tidak seimbang [10], [31].

$$Precision = \frac{TP}{TP + FP}, \qquad Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - Score = 2\,\frac{Precision * Recall}{Precision + Recall} \quad (5)$$

Di mana *TP*, *FP*, dan *FN* masing-masing menyatakan nilai *True Positive*, *False Positive*, dan *False Negative*.

Penelitian ini juga melakukan analisis statistik inferensial menggunakan uji *Chi-Square* yang bertujuan untuk mengetahui hubungan yang signifikan antara kategori sentimen dengan kategori *cognitive distortion* pada opini publik mengenai konflik anggaran pendidikan dan Program MBG. Hasil uji kemudian diinterpretasikan berdasarkan nilai p-*value* dengan tingkat signifikansi sebesar 0,05. Apabila nilai p-*value* lebih kecil dari 0,05, maka terdapat hubungan yang signifikan antara kedua variabel, sedangkan nilai yang lebih besar menunjukkan bahwa hubungan tersebut tidak signifikan. Seluruh hasil evaluasi dan analisis statistik digunakan sebagai dasar dalam menarik kesimpulan mengenai performa model klasifikasi sekaligus karakteristik hubungan antara sentimen publik dan *cognitive distortion* terhadap isu konflik anggaran pendidikan dan Program Makan Bergizi Gratis di Indonesia.

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (6)$$

$$V = \sqrt{\frac{\chi^2}{N\, min(r - 1, c - 1)}} \quad (7)$$

Di mana χ^2 menyatakan statistik uji *Chi-Square*, O_{ij} menyatakan frekuensi observasi riil pada baris i kolom j, E_{ij} menyatakan frekuensi harapan teoritis, V menyatakan koefisien asosiasi *Cramer's* V, N menyatakan ukuran sampel (2.000), r menyatakan jumlah baris tabel kontingensi, dan c menyatakan jumlah kolom tabel kontingensi.

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil penelitian yang mencakup enam aspek utama, yaitu karakteristik dataset hasil anotasi, reliabilitas antar-*rater*, sampel data hasil pemrosesan, kinerja klasifikasi kedua model, analisis efek kompresi probabilitas, serta korelasi statistik antara sentimen dan distorsi kognitif.

3.1. Karakteristik *Dataset*

Dataset penelitian ini terdiri dari 2.000 *tweet* opini publik berbahasa Indonesia terkait alokasi anggaran Makan Bergizi Gratis (MBG) dalam anggaran pendidikan pada APBN TA 2026. Data didistribusikan secara manual untuk dua lapis analisis (sentimen publik dan bias distorsi kognitif).

Tabel 1. Distribusi Sentimen Publik terhadap Isu Anggaran MBG

Sentimen	Jumlah *Tweet*	Persentase (%)
Negatif	1.651	82,5%
Netral	251	12,6%
Positif	98	4,9%
Total	2.000	100,0%

Berdasarkan Tabel 1, mayoritas sentimen publik berpolaritas Negatif (82,5%), yang secara langsung merepresentasikan masifnya kekhawatiran dan penolakan masyarakat terhadap inklusi dana operasional MBG pada pos anggaran pendidikan.

Tabel 2. Distribusi Label *Cognitive Distortion* (*CD*) pada *Dataset*

Tipe *Cognitive Distortion*	Jumlah *Tweet*	Persentase (%)
Mental Filter	163	8,2%
Catastrophizing	43	2,1%
Overgeneralization	36	1,8%
All-or-Nothing	34	1,7%
None (Tanpa Distorsi)	1.750	87,5%
Total *Dataset*	2.000	100,0%

Tabel 2 menunjukkan bahwa 250 opini (12,5%) dari total *dataset* memuat setidaknya satu jenis distorsi kognitif. *Mental Filter* teridentifikasi sebagai tipe yang paling dominan (8,2%). Akumulasi persentase pada tabel melampaui 100% karena penerapan pelabelan multi-label, yang memungkinkan satu *tweet* memiliki beberapa distorsi kognitif secara bersamaan.

3.2. Uji Reliabilitas Antar-*Rater*

Tabel 3 menyajikan hasil pengukuran kesepakatan *inter-rater* dari tiga anotator menggunakan *Fleiss' Kappa* (κ) pada 2.000 *tweet*.

Tabel 3. Koefisien Uji Reliabilitas *Fleiss' Kappa*

Kategori Label	Koefisien *Fleiss' Kappa* (κ)	Tingkat Kesepakatan (*Landis & Koch*)
Sentimen	0,9895	*Almost Perfect Agreement*
cd_catastrophizing	0,8945	*Almost Perfect Agreement*
cd_overgeneralization	0,8675	*Almost Perfect Agreement*

Kategori Label	Koefisien *Fleiss' Kappa* (κ)	Tingkat Kesepakatan (*Landis & Koch*)
*cd*_allornothing	0,9658	*Almost Perfect Agreement*
cd_mental_filter	0,8800	*Almost Perfect Agreement*
cd_none	0,9185	*Almost Perfect Agreement*

Berdasarkan Tabel 3, seluruh kategori pelabelan memperoleh nilai κ di atas 0,86. Menurut standar Landis & Koch, rentang nilai ini secara absolut masuk dalam kategori *Almost Perfect Agreement* (κ ≥ 0,81). Pencapaian ini memvalidasi tingginya kualitas *dataset* serta kuatnya kesamaan persepsi antar-*rater* dalam mengidentifikasi polaritas opini dan distorsi kognitif publik.

3.3. Data Hasil Pemrosesan

Subbab ini menyajikan sampel representatif dari tiga tahap pemrosesan data, yaitu hasil *preprocessing*, klasifikasi sentimen, dan deteksi distorsi kognitif, untuk memberikan transparansi kualitas data input model.

3.3.1. Data Hasil Preprocessing

Tabel 4. Sampel Data Hasil *Preprocessing*

No	Teks Mentah	Teks Bersih
1	Guru honorer gaji 300rb sebulan jomplang dengan pegawai MBG, ladang bisnis relawannya kah?	guru honorer gaji 300rb sebulan jomplang dengan pegawai makan bergizi gratis ladang bisnis relawannya kah
2	Soalnya utk kepentingan elektoral 2029 lbh menguntungkan MBG, Koperasi Merah Putih, Sekolah Rakyat drpd guru honorer.	soalnya untuk kepentingan elektoral 2029 lebih menguntungkan makan bergizi gratis koperasi merah putih sekolah rakyat daripada guru honorer

Pada Tabel 4, tahap *preprocessing* berhasil melakukan normalisasi kata tidak baku (misalnya "utk" menjadi "untuk", "lbh" menjadi "lebih") mengacu pada pemetaan bahasa gaul dan singkatan informal [23], menghapus tanda baca berlebih, serta mengonversi akronim "MBG" menjadi "makan bergizi gratis" untuk menyeragamkan representasi entitas program. Proses *stemming* sengaja tidak dilakukan agar imbuhan tetap utuh, mempertahankan nuansa emosi dan struktur semantik yang krusial bagi analisis sentimen dan pendeteksian distorsi kognitif [8], [15].

3.3.2. Data Hasil Klasifikasi Sentimen

Tabel 5. Sampel Data Hasil Klasifikasi Sentimen

No	Teks Bersih	Label Sentimen
1	guru honorer gaji 300rb sebulan jomplang dengan pegawai makan bergizi gratis ladang bisnis relawannya kah	Negatif
2	maka dari itu betapa pentingnya program makan bergizi gratis dimasukkan anggaran pendidikan sebesar 223 t agar tidak jadi negara gagal	Positif
3	kalau tidak mau di uraikan panjang jawab saja pertanyaan ini apakah benar sebagian anggaran pendidikan dialokasikan untuk makan bergizi gratis	Netral

Tweet pertama dikategorikan Negatif karena memuat kritik langsung terhadap disparitas kompensasi antara pegawai MBG dan guru honorer. *Tweet* kedua dikategorikan Positif karena

berisi dukungan kepada program. *Tweet* ketiga dikategorikan Netral karena berisi pertanyaan mengenai kebingungan warga X.

3.3.3. Data Hasil Deteksi Distorsi

Tabel 6. Sampel Data Hasil Deteksi Distorsi Kognitif

No	Teks Bersih	Label Sentimen	Label CD
1	guru honorer gaji 300rb sebulan jomplang dengan pegawai makan bergizi gratis ladang bisnis relawannya kah	Negatif	*Mental Filter*
2	makan bergizi gratis sangat berperan dalam pendidikan karakter terwujud manusia yang berkarakter menjilat menipu habis nirempati kapitalis mark up anggaran dan masih banyak lagi lanjutkan embege sampai kiamat	Negatif	*Catastrophizing, Mental Filter*
3	maka dari itu betapa pentingnya program makan bergizi gratis dimasukkan anggaran pendidikan sebesar 223 t agar tidak jadi negara gagal	Positif	*None*

Tabel 6 menunjukkan sifat multi-label pelabelan distorsi kognitif, di mana satu *tweet* dapat mengandung lebih dari satu tipe distorsi secara simultan. Seluruh *tweet* berdistorsi pada sampel ini berasal dari sentimen Negatif, sementara *tweet* Positif/Netral konsisten berlabel *None*; pola ini dianalisis lebih lanjut pada subbab korelasi statistik.

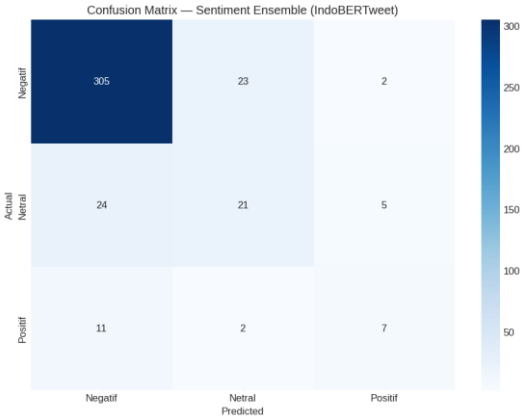
3.4. Evaluasi Kinerja Model Klasifikasi

Evaluasi dilakukan menggunakan *confusion matrix* dengan metrik *accuracy, precision, recall*, dan *F1-score* [10], [20]. Keterbatasan performa pada kelas Positif (*F1-score* 0,41) disebabkan oleh sedikitnya sampel positif (20 *tweet* pada data uji), yang merepresentasikan karakteristik riil wacana publik di media sosial X yang didominasi sentimen kritis.

Akurasi lapis pertama (83,25%) sejalan dengan Arunia dkk. [10] (83,00% dengan IndoBERT-Large) dan Mursyid dkk. [22] (80,11%). Dominasi sentimen negatif pada *dataset* ini (82,5%) jauh lebih ekstrem dibandingkan Arunia dkk. [10] (48,42%) maupun Sulistyo dan Setiadi [8] (68,6%), yang disebabkan oleh strategi kueri yang secara spesifik menjaring wacana konflik anggaran dan gugatan *Judicial Review*.

Tabel 7. Performa Per-Kelas Sentimen (Data Uji *Hold-out*)

Kelas Sentimen	*Precision*	*Recall*	*F1-Score*	*Support*
Negatif	0,90	0,92	0,91	330
Netral	0,46	0,42	0,44	50
Positif	0,50	0,35	0,41	20
Weighted Average	0,82	0,83	0,83	400
Macro Average	0,62	0,56	0,59	400

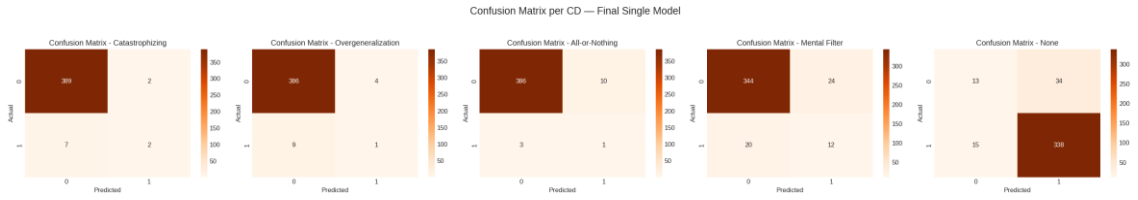


Gambar 1. *Confusion Matrix* Hasil Klasifikasi Sentimen

Klasifikasi distorsi kognitif pada lapis 2 menggunakan model tunggal IndoBERT-base dengan *threshold* 0,5, mencapai *Subset Accuracy* 83,75% dan *F1-macro* 37,19%. Hasil per-kelas disajikan pada Tabel 8. Meskipun *F1-macro* rendah akibat keterbatasan sampel positif pada kelas minoritas ekstrem, model tunggal dengan pembobotan kelas terbukti mampu melokalisasi sinyal deteksi tanpa bias kebocoran data.

Tabel 8. Performa Per-Kelas Distorsi Kognitif (Data Uji *Hold-out*)

Kategori Distorsi Kognitif	*Precision*	*Recall*	*F1-Score*	*Support*
Catastrophizing	0,50	0,22	0,31	9
Overgeneralization	0,20	0,10	0,13	10
All-or-Nothing	0,09	0,25	0,13	4
Mental Filter	0,33	0,38	0,35	32
None (Tanpa Distorsi)	0,91	0,96	0,93	353
Micro Average	0,83	0,87	0,85	408
Macro Average	0,41	0,38	0,37	408



Gambar 2. *Confusion Matrix* Hasil Klasifikasi Distorsi Kognitif

3.5. Analisis Efek Kompresi Probabilitas

Perbandingan empiris antara arsitektur model tunggal dan model *ensemble* menunjukkan bahwa model tunggal menghasilkan *F1-macro* lebih tinggi (0,3719 vs 0,3500). Efek *signal drowning* terlihat jelas pada kelas *Overgeneralization*: model *ensemble* mengalami kolaps total (*F1-score* 0,00) karena proses perataan probabilitas dari *fold* yang tidak melihat data positif menekan nilai di bawah ambang batas, sedangkan model tunggal berhasil mendeteksi kelas ini (*F1-score* 0,13). Peningkatan serupa terjadi pada *Catastrophizing* (0,31 vs 0,17). Temuan ini mengonfirmasi bahwa penggabungan *ensemble* tidak selalu optimal untuk klasifikasi multi-label dengan distribusi kelas minoritas yang ekstrem.

3.6. Analisis Korelasi Statistik Sentimen x Distorsi Kognitif

Untuk menguji signifikansi hubungan antara sentimen publik dan kecenderungan timbulnya distorsi kognitif, dilakukan tabulasi silang (Tabel 9) dan uji statistik *Chi-Square*, *Cramer's* V, serta *Fisher's Exact Test* (Tabel 10).

Tabel 9. Tabel Silang (*Crosstab*) Sentimen Publik dan Tipe CD

Tipe *Cognitive Distortion*	Negatif	Netral	Positif	Total
Mental Filter	163	0	0	163
Catastrophizing	43	0	0	43
Overgeneralization	35	1	0	36
All-or-Nothing	34	0	0	34

Tabel 10. Hasil Uji Statistik Korelasi dan Asosiasi

Variabel Distorsi Kognitif	*Chi-Square* (χ^2)	*Cramer's* V	p-*value* (*Chi-Square*)	p-*value* (*Fisher's Exact*)	Keterangan
Mental Filter	37,5134	0,1370	< 0,001	< 0,001	Sangat Signifikan
Catastrophizing	9,2894	0,0682	0,0096	< 0,001	Sangat Signifikan
All-or-Nothing	7,3115	0,0605	0,0258	0,0023	Sangat Signifikan
Overgeneralization	5,5419	0,0526	0,0626	0,0136	Signifikan

Berdasarkan Tabel 9 dan 10, seluruh tipe distorsi kognitif terkonsentrasi penuh pada sentimen Negatif. Karena *expected cell count* < 5 pada tabel kontingensi, dilakukan penggabungan kategori menjadi biner (Negatif vs Non-Negatif) dan diuji menggunakan *Fisher's Exact Test* dengan *Yates' Correction*. Hasil menunjukkan semua variabel CD berhubungan signifikan ($p < 0,05$) dengan sentimen negatif. *Mental Filter* memiliki asosiasi terkuat ($V = 0,1370$, $p < 0,001$), diikuti *Catastrophizing* ($V = 0,0682$, $p < 0,001$), *All-or-Nothing* ($V = 0,0605$, $p = 0,0023$), dan *Overgeneralization* ($p = 0,0136$).

3.7 Pembahasan

Penelitian ini berhasil menjawab ketiga tujuan yang ditetapkan. Pertama, distribusi sentimen menunjukkan penolakan masif (82,5% Negatif) terhadap penempatan anggaran MBG dalam pos pendidikan (Tabel 1), proporsi yang jauh melampaui studi MBG terdahulu [8], [9], [10]. Tingginya proporsi ini mengindikasikan bahwa *framing* isu sebagai konflik alokasi lintas sektor, alih-alih evaluasi program MBG secara mandiri, memperburuk *tone* opini publik karena wacana yang beredar di platform X lebih menyoroti pengurangan dana BOS dan gaji guru honorer dibandingkan manfaat program itu sendiri. Kedua, *Mental Filter* teridentifikasi sebagai distorsi kognitif paling dominan (8,2%, Tabel 2), diikuti *Catastrophizing*, *Overgeneralization*, dan *All-or-Nothing* dengan proporsi yang jauh lebih kecil. Ketiga, uji *Fisher's Exact* membuktikan hubungan signifikan ($p < 0,05$) seluruh tipe CD dengan sentimen negatif (Tabel 9 dan 10), dengan *Mental Filter* memiliki asosiasi terkuat ($V = 0,1370$, $p < 0,001$), diikuti *Catastrophizing* ($V = 0,0682$, $p < 0,001$), *All-or-Nothing* ($V = 0,0605$, $p = 0,0023$), dan *Overgeneralization* ($p = 0,0136$).

Konsentrasi distorsi kognitif pada sentimen negatif ini konsisten dengan kerangka teori CBT oleh Beck [7], yang mendefinisikan *Mental Filter* sebagai kecenderungan mengekstraksi detail negatif tunggal dan menjadikannya representasi keseluruhan situasi [15]. Publik yang terdistorsi cenderung hanya "menyaring" informasi pemangkasan BOS dan gaji guru honorer, tanpa mempertimbangkan argumentasi pemerintah mengenai dasar hukum Pasal 22 ayat (3) UU 17/2025 maupun mekanisme pengelolaan MBG oleh BGN. Pola serupa tampak pada *Catastrophizing*, di mana sebagian opini melebih-lebihkan dampak alokasi anggaran ini seolah-olah akan berujung pada kegagalan sistem pendidikan secara menyeluruh. Fenomena seleksi

informasi sepihak ini konsisten dengan review sistematis Xing dkk. [32] yang mengidentifikasi bahwa bias kognitif merupakan pendorong utama polarisasi opini di media sosial, serta efek *echo chamber* yang memperkuat narasi sepihak [33].

Perbedaan dengan Muhabbab dkk. [9] yang menemukan dominasi sentimen positif (47,2%) pada periode Januari hingga Agustus 2025 memperkuat argumen bahwa persepsi publik terhadap program MBG bersifat dinamis dan dipengaruhi konteks temporal, karena penelitian ini berfokus pada periode konflik anggaran (Januari hingga Juni 2026) yang eksplisit menyoroti kompetisi alokasi dana antar-sektor, bukan evaluasi manfaat gizi program secara langsung. Perbedaan tajam ini menegaskan bahwa strategi pengambilan sampel turut menentukan polaritas dominan yang teramati, sehingga perbandingan lintas studi MBG perlu mempertimbangkan periode dan konteks isu yang melatarbelakangi wacana publik yang dianalisis.

Keberhasilan *pipeline* dua lapis paralel memvalidasi prinsip penghindaran *negative transfer* [19], di mana pemisahan tugas klasifikasi sentimen (afektif) dan deteksi distorsi kognitif (penalaran) ke dalam dua model independen menghasilkan performa yang lebih stabil dibandingkan pendekatan gabungan (Tabel 7 dan 8). Temuan bahwa model tunggal lebih unggul dari *ensemble* pada kelas minoritas ekstrem (*F1-macro* 0,3719 vs 0,3500) merupakan kontribusi yang belum banyak dibahas dalam literatur deteksi distorsi kognitif komputasional [11], [12], [13], [14]. Pencapaian *F1-macro* 0,3719 pada *dataset* dengan hanya 12,5% data CD perlu dikontraskan dengan Sastra dkk. [15] yang memiliki proporsi lebih seimbang (51,8%), menegaskan bahwa tantangan deteksi CD pada teks kebijakan publik bersifat unik dibandingkan domain klinis.

Penelitian ini memberikan kontribusi dari tiga perspektif. Pertama, dari perspektif *Computational Social Science*, penelitian ini merupakan upaya awal menerapkan deteksi CD pada domain kebijakan fiskal [11], [12], [13], [14], memperluas cakupan aplikasi deteksi distorsi kognitif yang sebelumnya terbatas pada domain klinis dan kesehatan mental ke ranah wacana publik non-klinis. Kedua, dari perspektif metodologis, arsitektur *pipeline* paralel dual-BERT memberikan *blueprint* yang dapat direplikasi untuk kasus serupa di mana dua tugas klasifikasi memiliki dasar linguistik berbeda namun berasal dari sumber data yang sama. Ketiga, dari perspektif kebijakan publik, korelasi signifikan antara sentimen negatif dan *Mental Filter* memberikan dasar empiris bagi strategi komunikasi pemerintah yang lebih transparan [33], [32], khususnya dalam mengklarifikasi dasar hukum dan mekanisme alokasi anggaran MBG kepada publik.

Keterbatasan penelitian ini meliputi: (1) data hanya bersumber dari platform X, sehingga belum merepresentasikan sentimen publik lintas platform; (2) jumlah sampel kelas minoritas sangat terbatas, yang merupakan tantangan umum pada klasifikasi multi-label dengan distribusi tidak seimbang [31]; dan (3) taksonomi distorsi kognitif dibatasi empat tipe dari sepuluh yang ada [15], sehingga pola distorsi lain belum tercakup dalam analisis ini. Penelitian selanjutnya direkomendasikan untuk memperluas sumber data lintas platform, memanfaatkan teknik augmentasi data atau *resampling* yang dirancang khusus untuk kelas minoritas multi-label[31], serta menerapkan pendekatan LLM [12], [14].

4. Simpulan

Penelitian ini berhasil menerapkan arsitektur *pipeline* NLP dua lapis paralel untuk menganalisis sentimen publik dan mendeteksi distorsi kognitif pada wacana anggaran MBG dalam APBN 2026. *Mental Filter* teridentifikasi sebagai distorsi paling dominan (8,2%), diikuti *Catastrophizing* (2,1%), *Overgeneralization* (1,8%), dan *All-or-Nothing* (1,7%). Lapis pertama (sentimen, IndoBERTweet, *5-fold ensemble*) mencapai akurasi 83,25%, sedangkan lapis kedua (CD, IndoBERT-base, model tunggal) mencapai *subset accuracy* 83,75%. Model tunggal terbukti lebih unggul dari *ensemble* pada kelas minoritas ekstrem (*F1-score Overgeneralization* meningkat dari 0,00 menjadi 0,13). Uji *Fisher's Exact* membuktikan hubungan signifikan ($p < 0,05$) antara sentimen negatif dan seluruh tipe CD, terutama *Mental Filter* ($p < 0,001$). Keterbatasan meliputi fokus pada satu platform dan jumlah sampel kelas minoritas yang sangat terbatas. Penelitian selanjutnya direkomendasikan untuk memperluas sumber data dan memanfaatkan LLM untuk meningkatkan sensitivitas klasifikasi pada kelas langka.

Daftar Referensi

[1] "UU 17 TAHUN 2025 | Anggaran Pendapatan dan Belanja Negara Tahun Anggaran 2026." Accessed: Jul. 12, 2026. [Online]. Available: https://djpb.kemenkeu.go.id/kppn/jakarta3/id/download/peraturan-terbaru/3051-uu-17-tahun-2025-anggaran-pendapatan-dan-belanja-negara-tahun-anggaran-2026.html

[2] "Direktorat Jenderal Perimbangan Keuangan | Peraturan Presiden Nomor 118 Tahun 2025 tentang Rincian Anggaran Pendapatan dan Belanja Negara Tahun Anggaran 2026." Accessed: Jul. 12, 2026. [Online]. Available: https://djpk.kemenkeu.go.id/?p=72421

[3] "Perpres No. 115 Tahun 2025." Accessed: Jul. 12, 2026. [Online]. Available: https://peraturan.bpk.go.id/Details/343430/perpres-no-115-tahun-2025

[4] "Permohonan Pengujian Undang-Undang Nomor 17 Tahun 2025 tentang APBN Tahun Anggaran 2026 Terhadap UUD 1945 Perkara No. 40, 52, dan 55/PUU-XXIV/2026." Accessed: Jul. 12, 2026. [Online]. Available: https://tracking.mkri.id/index.php?page=web.TrackPerkara&id=52/PUU-XXIV/2026

[5] P. H. Nehe, S. S. Berutu, and H. Budiati, "Analisis Sentimen Opini Masyarakat Terhadap Presiden Jokowi Sebelum Dan Sesudah Pilpres 2024 Menggunakan Metode Naive Bayes Classification," *Jutisi : Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 13, no. 1, p. 451, Apr. 2024, doi: 10.35889/jutisi.v13i1.1841.

[6] K. Syasyila, L. L. Gin, H. Abdullah Mohd. Nor, and M. R. Kamaluddin, "The role of cognitive distortion in criminal behavior: a systematic literature review," *BMC Psychol.*, vol. 12, no. 1, p. 741, Dec. 2024, doi: 10.1186/s40359-024-02228-0.

[7] Beck and Judith S, "Cognitive Behavior Therapy: Basics and Beyond (Third Edition)," New York, 2021.

[8] D. A. Sulistyo and E. Setiadi, "Analisis Sentimen Kebijakan Makan Bergizi Gratis Menggunakan IndoBERT dan Machine Learning," *JURNAL FASILKOM*, vol. 15, no. 3, pp. 507–513, Dec. 2025, doi: 10.37859/jf.v15i3.10546.

[9] A. Z. Muhabbab, B. Bunyamin, and H. Hasmawati, "Sentiment Analysis of Indonesia's Free Nutritious Meal Program on Platform X (Formerly Twitter) Using IndoBERT," *ZERO: Jurnal Sains, Matematika dan Terapan*, vol. 9, no. 3, p. 884, Dec. 2025, doi: 10.30829/zero.v9i3.27629.

[10] A. P. Arunia, R. R. Sani, I. N. Dewi, and M. T. Sulistyono, "Exploring Public Opinion on the 'Makan Bergizi Gratis' Program on X: A Comparative Analysis of IndoBERT-Large and NusaBERT-Large Models," *Journal of Applied Informatics and Computing*, vol. 10, no. 1, pp. 856–864, Feb. 2026, doi: 10.30871/jaic.v10i1.11757.

[11] S. Shreevastava and P. Foltz, "Detecting Cognitive Distortions from Patient-Therapist Interactions," in *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 151–158. doi: 10.18653/v1/2021.clpsych-1.17.

[12] Z. Chen, Y. Lu, and W. Wang, "*Empowering Psychotherapy with Large Language Models: Cognitive Distortion Detection through Diagnosis of Thought Prompting*," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 4295–4304. doi: 10.18653/v1/2023.findings-emnlp.284.

[13] S. Lin, Y. Wang, J. Dong, and S. Ni, "Detection and Positive Reconstruction of Cognitive Distortion Sentences: Mandarin Dataset and Evaluation," in *Findings of the Association for Computational Linguistics ACL 2024*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2024, pp. 6686–6701. doi: 10.18653/v1/2024.findings-acl.399.

[14] S. Lim, Y. Kim, C.-H. Choi, J. Sohn, and B.-H. Kim, "ERD: A Framework for Improving LLM Reasoning for Cognitive Distortion Classification," Mar. 2024.

[15] N. P. Sastra, Linawati, G. Sukadarmika, I. P. G. H. Suputra, N. M. Ariwilani, and N. L. I. D. Swandi, "Dataset on cognitive distortions for text classification in Indonesian

language," *Data Brief*, vol. 61, p. 111836, Aug. 2025, doi: 10.1016/j.dib.2025.111836.

[16] G. I. Winata *et al.*, "NusaX: Multilingual Parallel Sentiment Dataset for 10 Indonesian Local Languages," in *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 815–834. doi: 10.18653/v1/2023.eacl-main.57.

[17] S. Cahyawijaya *et al.*, "NusaCrowd: Open Source Initiative for Indonesian NLP Resources," in *Findings of the Association for Computational Linguistics: ACL 2023*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 13745–13818. doi: 10.18653/v1/2023.findings-acl.868.

[18] F. Koto, J. H. Lau, and T. Baldwin, "IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 10660–10668. doi: 10.18653/v1/2021.emnlp-main.833.

[19] W. Zhang, L. Deng, L. Zhang, and D. Wu, "A Survey on Negative Transfer," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 2, pp. 305–329, Feb. 2023, doi: 10.1109/JAS.2022.106004.

[20] N. K. Nissa and E. Yulianti, "Multi-label text classification of Indonesian customer reviews using bidirectional encoder representations from transformers language model," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 5, p. 5641, Oct. 2023, doi: 10.11591/ijece.v13i5.pp5641-5652.

[21] Q. Chen, J. Du, A. Allot, and Z. Lu, "LitMC-BERT: Transformer-Based Multi-Label Classification of Biomedical Literature With An Application on COVID-19 Literature Curation," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 19, no. 5, pp. 2584–2595, Sep. 2022, doi: 10.1109/TCBB.2022.3173562.

[22] Muhammad Mursyid, A. Setiawan, and M. Arifin, "Implementation of IndoBERT in Sentiment Analysis of Free Nutritious Meal Programs on the X Social Media Platform," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 19, no. 1, pp. 1228–1242, Mar. 2026, doi: 10.24036/jtip.v19i1.1104.

[23] R. Budianoor, S. W. Saputro, F. Abadi, R. A. Nugroho, and A. Farmadi, "Quantifying the Impact of Text Preprocessing on IndoBERT Fine-Tuning for Indonesian Informal Culinary Sentiment Analysis," *Journal of Computing Theories and Applications*, vol. 3, no. 4, pp. 564–581, May 2026, doi: 10.62411/jcta.15980.

[24] K. Martskvishvili, M. Panjikidze, T. Kamushadze, N. Garuchava, and D. J. A. Dozois, "Measuring the Thinking Styles: Psychometric Properties of the Georgian Version of the Cognitive Distortion Scale," *J. Cogn. Psychother.*, vol. 35, no. 4, pp. 268–289, Nov. 2021, doi: 10.1891/JCPSY-D-19-00025.

[25] F. Moons and E. Vandervieren, "Measuring agreement among several raters classifying subjects into one or more (hierarchical) categories: A generalization of Fleiss' kappa," *Behav. Res. Methods*, vol. 57, no. 10, p. 287, Sep. 2025, doi: 10.3758/s13428-025-02746-8.

[26] K. L. Gwet, *Handbook of Inter-Rater Reliability Fith Edition The Definitive Guide to Measuring the Extent of Agreement Among Raters Volume 1 Analysis of Categorical Ratings*, Fifth Edition. AgreeStat Analytics, 2021. [Online]. Available: https://sites.fastspring.com/agreestat/instant/cac5ed978_1_7923_5463_2ehttps://agreestat.com/books/cac5/https://agreestat.com/books/

[27] A. N. Uma, T. Fornaciari, D. Hovy, S. Paun, B. Plank, and M. Poesio, "Learning from Disagreement: A Survey," *Journal of Artificial Intelligence Research*, vol. 72, pp. 1385–1470, Dec. 2021, doi: 10.1613/jair.1.12752.

[28] A. Fatwanto, F. Zamakhsyari, and R. Ndungi, "A Systematic Literature Review of BERT-based Models for Natural Language Processing Tasks," *JURNAL INFOTEL*, vol. 16, no. 4, pp. 713–728, Dec. 2024, doi: 10.20895/infotel.v16i4.1206.

[29] N. Raghunathan and K. Saravanakumar, “Challenges and Issues in Sentiment Analysis: A Comprehensive Survey,” *IEEE Access*, vol. 11, pp. 69626–69642, 2023, doi: 10.1109/ACCESS.2023.3293041.

[30] F. Indriani, R. A. Nugroho, M. R. Faisal, and D. Kartini, “Comparative Evaluation of IndoBERT, IndoBERTweet, and mBERT for Multilabel Student Feedback Classification,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 6, pp. 748–757, Dec. 2024, doi: 10.29207/resti.v8i6.6100.

[31] A. N. Tarekegn, M. Giacobini, and K. Michalak, “A review of methods for imbalanced multi-label classification,” *Pattern Recognit.*, vol. 118, p. 107965, Oct. 2021, doi: 10.1016/j.patcog.2021.107965.

[32] Y. Xing, J. Z. Zhang, V. C. Storey, and A. Koohang, “Diving into the divide: a systematic review of cognitive bias-based polarization on social media,” *Journal of Enterprise Information Management*, vol. 37, no. 1, pp. 259–287, Feb. 2024, doi: 10.1108/JEIM-09-2023-0459.

[33] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, “The echo chamber effect on social media,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 9, Mar. 2021, doi: 10.1073/pnas.2023301118.