Analisis Sentimen, Linguistik, dan Spasial Pada Program Makan Bergizi Gratis Menggunakan LSTM dan GRU

Sentiment, Linguistic, and Spatial Analysis of the Free Nutritious Meal Program Using LSTM and GRU

Dewi Asiah Shofiana[1], **Rhalasya Eleina Putri**[*2], **Rahman Taufik**[3], **Favorisen Rosyking Lumbanraja**[4]

Ilmu Komputer, Universitas Lampung, Indonesia

*E-mail : dewi.asiah@fmipa.unila.ac.id[1], elenrhalsya@gmail.com[*2], rahman.taufik@fmipa.unila.ac.id[3], favorisen.lumbanraja@fmipa.unila.ac.id[4]*

**Corresponding author*

Received 23 May 2026; Revised 19 June 2026; Accepted 1 July 2026

Abstrak - Penelitian ini memetakan sentimen publik serta pola linguistik, temporal, dan spasial opini mengenai Program Makan Bergizi Gratis (MBG) pada X, sekaligus membandingkan kinerja Long Short-Term Memory (LSTM) dan Gated Recurrent Unit (GRU) untuk klasifikasi sentimen. Penelitian menggunakan pendekatan kuantitatif berbasis text mining pada tweet berbahasa Indonesia periode Januari sampai Agustus 2025. Pengumpulan data berbasis kata kunci menghasilkan 13.486 tweet yang difokuskan pada konteks sekolah menengah atas dan menjadi 12.476 tweet setelah penghapusan duplikasi. Pelabelan dilakukan melalui anotasi manual dan bantuan IndoBERT, sedangkan ketidakseimbangan kelas pada data latih ditangani menggunakan ADASYN. Evaluasi dilakukan melalui stratified 10-fold cross-validation dan data uji terpisah. LSTM memberikan kinerja terbaik dengan akurasi 77,63%, presisi 77,51%, recall 77,40%, F1-score 77,46%, dan AUC 91,54%, sedangkan GRU memperoleh akurasi 75,59% dan F1-score 75,40%. Analisis linguistik menunjukkan bahwa sentimen negatif berpusat pada kualitas makanan dan anggaran, sedangkan sentimen positif menonjolkan manfaat gizi, pemerataan, dan masa depan anak. Temuan ini menegaskan bahwa evaluasi kebijakan perlu menggabungkan performa model dengan konteks linguistik dan wilayah.

Kata kunci: Analisis Sentimen, Gated Recurrent Unit, Long Short-Term Memory, Pemrosesan Bahasa Alami, Program Makan Bergizi Gratis

Abstract - *This study maps public sentiment and linguistic, temporal, and spatial patterns in opinions about the Free Nutritious Meals Program (MBG) on X and compares Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) for sentiment classification. A quantitative text-mining design was applied to Indonesian-language tweets posted from January to August 2025. Keyword-based crawling produced 13,486 tweets, which were filtered for senior high school contexts and reduced to 12,476 after duplicate removal. Sentiment labels were assigned through manual annotation and IndoBERT-assisted labeling, while class imbalance in the training set was handled using ADASYN. Model performance was evaluated using stratified 10-fold cross-validation and a separated test set. LSTM achieved the best test performance with 77.63% accuracy, 77.51% precision, 77.40% recall, 77.46% F1-score, and 91.54% AUC, while GRU obtained 75.59% accuracy and 75.40% F1-score. Linguistic exploration showed that negative sentiment centered on food quality and budget issues, whereas positive sentiment emphasized nutrition, equity, and children's future. These findings show that policy evaluation should combine model performance with linguistic and regional context.*

Keywords: *Free Nutritious Meals Program, Gated Recurrent Unit, Long Short-Term Memory, Natural Language Processing, Sentiment Analyst*

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

1. PENDAHULUAN

Media sosial kini menjadi ruang utama bagi publik untuk menyampaikan keluhan, dukungan, maupun kritik terhadap isu-isu terkini. Salah satu kebijakan yang paling banyak diperbincangkan di *platform* X adalah Program Makan Bergizi Gratis (MBG). Program ini merupakan inisiatif utama pemerintah yang bertujuan menyediakan makan siang gratis bagi anak-anak sekolah sebagai langkah preventif terhadap kasus stunting di Indonesia. Meski niatnya baik, program ini memicu pro dan kontra yang sangat luas di masyarakat. Kekhawatiran publik diawali pada besarnya alokasi anggaran dan efisiensi pelaksanaannya di lapangan. Pada implementasinya juga memunculkan perdebatan mengenai kualitas layanan, kesiapan pelaksana, dan arah prioritas kebijakan [1]. Sayangnya, memantau sentimen ini secara manual tidak lagi memungkinkan. Algoritma media sosial cenderung menciptakan *filter bubble* [2] dan *echo chamber* [3], jadi pengguna hanya akan terpapar pada opini yang sejalan dengan pandangan pribadinya. Akibatnya, observasi biasa akan menghasilkan kesimpulan yang bias dan parsial. Untuk mengatasi batasan tersebut, teknologi kecerdasan buatan khususnya di bidang *Natural Language Processing* (NLP) menjadi solusi yang sangat relevan.

Pendekatan *deep learning* mampu memproses ribuan data teks secara otomatis dan memahami konteks bahasa manusia yang rumit. Pendekatan mesin pembelajaran konvensional memiliki kelemahan bawaan, yakni ketidakmampuannya memahami urutan dan relasi makna antarkata dalam sebuah kalimat utuh. Untuk menangani data teks cuitan yang bersifat sekuensial dan sangat dinamis, arsitektur deep learning berbasis *Recurrent Neural Network* (RNN) terbukti jauh lebih relevan. Diantara berbagai arsitektur jaringan saraf tiruan, model *Long Short-Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU) dikenal tangguh dalam menangani data teks yang bersifat sekuensial [4]. Kedua model ini dipilih untuk mengatasi kelemahan pada jaringan saraf tradisional dengan kemampuannya menyimpan memori jangka panjang dari sebuah kalimat karena keduanya mampu mempelajari konteks antarkata secara bertahap [5] yang ditunjukan dengan mekanisme *gating* pada kedua model tersebut. Pada LSTM dikenal sangat optimal dalam mengenali data teks di media sosial, sedangkan pada GRU menawarkan keunggulan berupa arsitektur yang lebih sederhana sehingga proses komputasinya menjadi lebih cepat namun akurat [6].

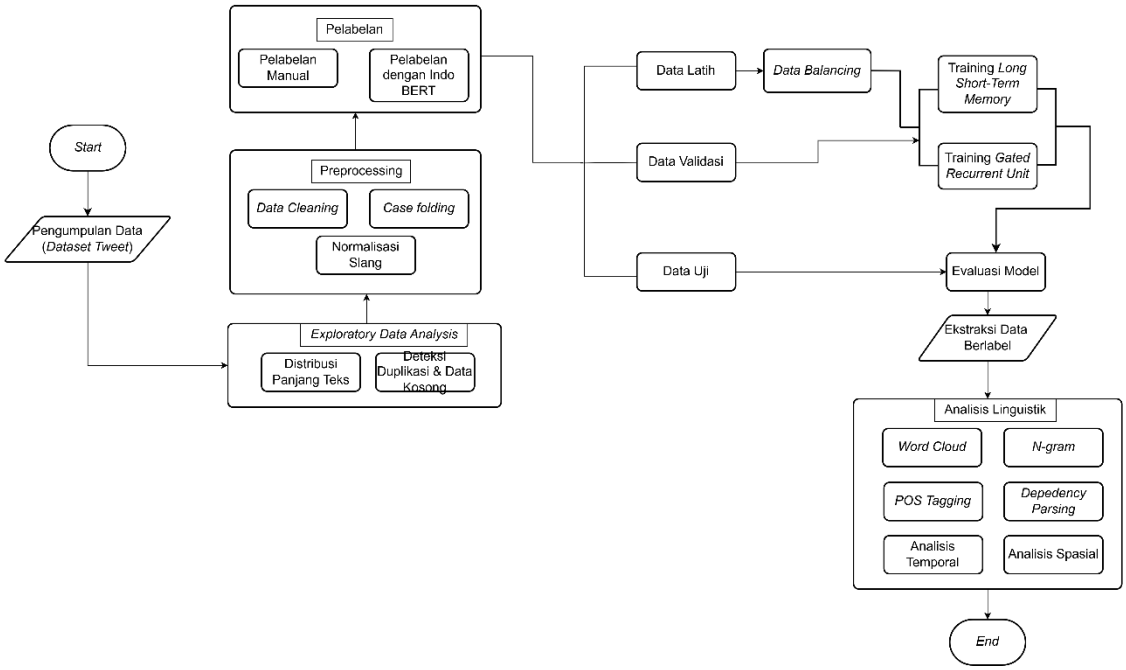
Sebagian besar penelitian terdahulu masih berfokus pada performa model tanpa mengkaji lebih dalam pola linguistik. Penelitian terdahulu pada isu yang sama dengan algoritma SVM berbasis TF-IDF melaporkan akurasi 77,5% [7], Indo-BERT dan Bi-LSTM sekitar 80% dan 78% [8], Bi-GRU 80% [9], sedangkan *Random Forest* bahkan mencapai lebih dari 90% pada korpus yang berbeda [10]. Perbedaan ini menegaskan bahwa kinerja model sangat dipengaruhi oleh karakter data, skema pelabelan, serta strategi penyeimbangan kelas, sehingga angka performa tidak dapat dibaca terlepas dari konteks korpus. Penelitian ini melakukan dua hal utama, pertama membandingkan kinerja LSTM dan GRU pada analisis sentimen tiga kelas terhadap *tweet* berbahasa Indonesia tentang program MBG. Selanjutnya, menganalisis secara tekstual pada hasil klasifikasi dengan analisis n-gram, POS *tagging*, *dependency parsing*, analisis dinamika temporal, dan analisis spasial antarpulau. Hal tersebut menunjukkan model yang lebih baik, tetapi juga menjelaskan isu apa yang mendorong sentimen serta bagaimana sentimen berubah menurut waktu dan wilayah.

Kajian sentimen terhadap kebijakan publik tidak cukup berhenti pada klasifikasi sentimen positif, negatif, dan netral. Pembacaan sentimen perlu dipadukan dengan analisis linguistik agar alasan dibalik setiap label dapat dipahami secara substantif [11]. Penelitian ini dilakukan sebagai upaya membaca opini publik secara lebih utuh dan berupaya menjelaskan terkait model terbaik dalam menganalisis sentimen, serta juga bagaimana wacana kebijakan dibentuk, diperdebatkan, dan didistribusikan dalam ruang digital. Melalui pendekatan menyeluruh ini, penelitian diharapkan mampu memberikan wawasan mendalam bagi pemerintah mengenai faktor utama pendorong opini masyarakat sekaligus membuktikan keandalan arsitektur *deep learning* dalam pemrosesan teks berbahasa Indonesia.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

2. METODE PENELITIAN

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap Program MBG menggunakan algoritma LSTM dan GRU yang dipadukan dengan ADASYN. Tahapan penelitian dimulai dari pengumpulan data melalui proses *crawling*, dilanjutkan dengan tahapan *exploratory data analysis* untuk melihat karakteristik data, kemudian *preprocessing* untuk membersihkan dan menyamakan format data teks. Pada tahapan pelabelan, dilakukan pelabelan manual yang dikombinasikan dengan pelabelan IndoBERT. Hal tersebut diakibatkan oleh data terindikasi yang tidak seimbang, teknik ADASYN diterapkan pada data latih. Model kemudian dilatih dan dievaluasi kinerjanya menggunakan berbagai metrik. Pada tahap akhir, hasil prediksi dibedah lebih lanjut menggunakan analisis linguistik dan spasial-temporal. Tahapan-tahapan tersebut ditunjukan pada Gambar 1 yang merupakan alur penelitian.



Gambar 1. Diagram Alir Penelitian

2.1 Pengumpulan Data

Data penelitian dikumpulkan dari media sosial X menggunakan teknik *crawling* dengan *tools tweet harvest* milik [12]. Pada penelitian ini, *crawling data* dilakukan dengan kata kunci “makan bergizi gratis”, “MBG", “program gizi siswa", dan “program makan Prabowo”. Data yang terkumpul kemudian difilter dengan kata kunci yang memfokuskan *case study* di SMA. Diterapkan kata kunci “SMA”, “SMK”, “SMAN”, “SMAS”, “Sekolah Menengah Atas”. Pengumpulan data dilakukan dari periode Januari 2025 – Agustus 2025 yang berjumlah 13.486 tweet dengan variabel created_at, full_text, dan label dengan format CSV.

2.2 *Exploratory Data Analysis*

Pada tahapan EDA dilakukan tiga aspek utama yaitu, distribusi panjang kata, distribusi panjang karakter, dan deteksi duplikasi *tweet*. Tahapan ini penting karena memberikan gambaran awal mengenai variasi bentuk teks, kompleksitas kalimat, serta potensi adanya data berulang yang dapat memengaruhi performa model. Berdasarkan hasil EDA diketahui bahwa panjang kata maksimum pada dataset adalah 67 kata. Hal ini mengindikasikan bahwa sebagian besar *tweet* disampaikan dalam bentuk opini yang relatif pendek namun tetap memuat informasi yang cukup.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

Pada tahapan deteksi duplikasi *tweet*, ditemukan 1.010 data yang identik sehingga jumlah data berkurang menjadi 12.476 *tweet*.

2.3 *Preprocessing Data*

Preprocessing meliputi *case folding*, *cleaning*, dan normalisasi kata tidak baku agar representasi teks lebih konsisten. *Preprocessing data* adalah tahapan penting sebelum proses pelabelan dan pemodelan. Tujuannya untuk memastikan bahwa teks yang akan dijadikan masukan bagi model bersih, konsisten, dan tidak mengandung duplikat yang dapat mempengaruhi efektivitas algoritma.

2.3.1 *Case Folding*

Dilakukan proses *case folding* yaitu menyelaraskan huruf secara merata pada sebuah dokumen dengan mengubah semua huruf kapital menjadi huruf kecil seperti yang ditunjukan pada Tabel 1.

Tabel 1 Tahap *Case Folding*

Sebelum *Case Folding*	Sesudah *Case Folding*
@kangsolci kemewahan ga si MBG untuk anak SMA dpt ayam satu potong kek gt wkwkwk	@kangsolci kemewahan ga si mbg untuk anak sma dpt ayam satu potong kek gt wkwkwk

2.3.2 *Cleaning*

Cleaning adalah proses pembersihan dilakukan untuk menghilangkan unsur yang tidak relevan seperti URL, penyebutan (@namapengguna), tagar, dan simbol non-alfabet seperti yang ditunjukan Tabel 2.

Tabel 2 Tahap *Cleaning*

Sebelum *Cleaning*	Sesudah *Cleaning*
@kangsolci kemewahan ga si mbg untuk anak sma dpt ayam satu potong kek gt wkwkwk	kemewahan ga si mbg untuk anak sma dpt ayam satu potong kek gt wkwkwk

2.3.3 Normalisasi

Normalisasi adalah proses mengubah *tweet* yang menggunakan bahasa gaul atau singkatan menjadi kata baku sesuai kamus berdasarkan penelitian sebelumnya milik [13], [14] seperti pada Tabel 3.

Tabel 3 Tahap Normalisasi

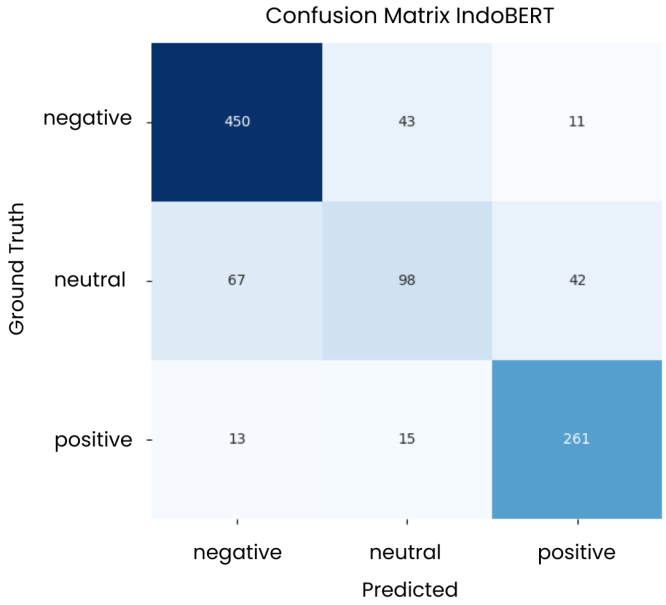
Sebelum Normalisasi	Sesudah Normalisasi
kemewahan ga si mbg untuk anak sma dpt ayam satu potong kek gt wkwkwk	kemewahan tidak sih mbg untuk anak sma dapat ayam satu potong seperti begitu wkwkwk

2.4 Pelabelan

Dilakukan penggabungan metode pelabelan manual dan pelabelan IndoBERT dikarenakan butuhnya pelabelan yang valid namun tetap cepat dan efisien. Pada pelabelan manual dilakukan pada 1000 data yang didasarkan pada kamus sentimen penelitian terdahulu milik [15] dan [16] yang dilakukan dengan 3 anotator dengan sentimen negatif, netral, dan positif. Kemudian, pada sisa data akan dilabeli secara otomatis dengan IndoBERT dengan menjadikan 1000 data sebelumnya sebagai acuan. Hal ini umum dilakukan oleh penelitian-penelitian terkait analisis sentimen [17]. Untuk menjamin validitas *ground truth* (data acuan) yang dihasilkan,

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

dilakukan dua tahap pengujian. Pertama, dilakukan evaluasi kinerja IndoBERT dalam melabeli 1.000 data acuan tersebut. Hasil evaluasi menunjukan bahwa IndoBERT mencapai akurasi sebesar 80,9% dan *f1-score* sebesar 80,06% dalam mereplikasi label manual, yang mengindikasikan bahwa model mampu mempelajari pola pelabelan dengan baik sebelum diterapkan pada keseluruhan *dataset*. Distribusi kesalahan klasifikasi antar kelas sentimen ditunjukkan pada *confusion matrix* pada Gambar 2.



Gambar 2. *Confusion Matrix* IndoBERT

Kedua, untuk menguji reliabilitas kesepakatan antar anotator dalam proses pelabelan manual dilakukan perhitungan koefisien Fleiss' Kappa pada 150 data yang dilabeli secara independen oleh ketiga annotator. Sampel sebanyak 150 data dipilih menggunakan teknik *stratified sampling* agar proporsi setiap kelas sentiment tetap representatif terhadap keseluruhan data acuan. Pendekatan ini umum digunakan dalam pengujian reliabilitas antar anotator (*inter-annotator agreement*) untuk menjaga efisiensi proses anotasi tanpa mengurangi validitas hasil pengukuran kesepakatan Fleiss' Kappa dipilih karena penelitian ini melibatkan 3 anotator dalam proses pelabelan manual, sedangkan metode pengukuran kesepakatan yang lebih umum seperti Cohen's Kappa hanya dirancang untuk membandingkan kesepakatan antara dua anotator. Fleiss' Kappa memungkinkan perhitungan tingkat kesepakatan dari ketiga anotator tersebut secara bersamaan dalam satu perhitungan. Selain itu, metode ini juga memperhitungkan kemungkinan kesepakatan yang terjadi secara kebetulan (*chance agreement*), sehingga hasil yang didapatkan lebih merepresentasikan kesepakatan substantif antar anotator, bukan sekadar kecocokan acak. Nilai Kappa dihitung menggunakan persamaan berikut [18].

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

$\bar{P}$ merupakan rata-rata proporsi kesepakatan yang teramati (*observed agreement*) diantara seluruh annotator pada setiap data, dan $\bar{P}_e$ merupakan proporsi kesepakatan yang diharapkan terjadi secara kebetulan (*expected agreement by chance*) yang dihitung berdasarkan distribusi proporsi setiap kategori label dari seluruh anotator. Hasil pengujian menunjukan nilai Kappa sebesar 83,42% yang menurut kriteria pada penelitian Landis dan Koch [19] termasuk dalam kategori kesepakatan hampir sempurna (*almost perfect agreement*). Nilai ini menunjukan bahwa ketiga anotator memiliki tingkat kesepahaman yang tinggi dalam menentukan label sentimen, sehingga proses pelabelan manual yang digunakan sebagai acuan dalam penelitian ini dapat dinyatakan teruji. Berdasarkan hasil pelabelan, didapatkan 5.039 kelas negatif, 3.754 kelas netral, dan 3.682 kelas positif yang dapat dilihat contoh pelabelannya pada Tabel 4.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

Tabel 4 Tahap Pelabelan

Sentimen	*Tweet*
Negatif	@khalchive Mbg di sma saya, kenyang kaga bergizi jg kaya engga cm seuprit gt sayurnya
Netral	@HeryIdris5 Kandungan Gizi Belalang dan Ulat Sagu Serangga yang Berpotensi Jadi Menu MBG SMP, SMA https://t.co/v41il8Yl1G
Positif	Kebijakan yg tepat karena MBG untuk anak sekolahan khususnya SMA itu targetnya lebih luas dan lebih bermanfaat sehingga banyak rakyat Indonesia merasakan manfaatnya.

2.5 Pembagian Data dan *Data Balancing*

Penelitian ini menggunakan *Stratified K-Fold Cross Validation* (SKCV) untuk memperoleh performa model yang lebih reliabel dan mengurangi ketergantungan pada satu kali pemisahan data. Pada metode ini, seluruh data dibagi menjadi 10 *fold* yang proporsi label di dalam tiap *fold*-nya dipertahankan (*stratified*) agar mendekati distribusi label pada keseluruhan *dataset* [20]. Setelah tokenisasi, data dibagi secara terstratifikasi ke dalam data latih dan data uji. Ketidakseimbangan kelas pada data latih diatasi menggunakan *Adaptive Synthetic Sampling* (ADASYN). Data dibagi secara terstratifikasi menjadi 80% data latih dan 20% data uji agar evaluasi akhir dilakukan pada korpus yang benar-benar terpisah. Pada data latih, distribusi awal terdiri atas 4.031 *tweet* negatif, 3.003 netral, dan 2.946 positif. Penerapan ADASYN meningkatkan kelas netral menjadi 3.465 dan kelas positif menjadi 3.886, sementara kelas negatif dipertahankan 4.031. Strategi ini dipakai untuk mengurangi bias model terhadap kelas mayoritas tanpa mencemari distribusi alami pada data validasi dan data uji [21].

2.6 *Training Model*

Untuk menemukan konfigurasi hyperparameter yang paling optimal bagi kedua model, pencarian dilakukan menggunakan metode *Tree-structured Parzen Estimator* (TPE) melalui *library* Optuna. TPE dipilih karena algoritma optimasi berbasis probabilitas ini terbukti lebih efisien dan terarah dalam mengeksplorasi ruang pencarian hiperparameter yang kompleks dibandingkan metode pencarian acak (*random search*) konvensional [22].

Tabel 5 Konfigurasi Terbaik dengan Optuna TPE

Hyperparameter	LSTM	GRU
Embedding dimension	100	128
Spatial dropout	0.4	0.2
Dropout	0.4	0.5
Recurrent dropout	0.4	0.3
Dense dropout	0.5	0.3
L2 regularization	1e-4	1e-4
Learning rate	5e-4	1e-4
Clipnorm	1.0	1.0
Units	64	64

Dalam konfigurasi tersebut, penggunaan *SpatialDropout1D*, *dropout*, dan *recurrent dropout* diimplementasikan sebagai teknik regularisasi struktural untuk mencegah model terlalu menghafal data latih dengan cara menonaktifkan sebagian neuron secara acak [23]. Regularisasi L2 juga ditambahkan untuk mengontrol besaran bobot jaringan agar model lebih general. Selain itu, *optimizer* Adam dipilih karena kemampuannya dalam melakukan adaptasi *learning rate* secara dinamis, yang dipadukan dengan teknik *clipnorm* guna mencegah terjadinya fenomena *exploding gradient* yang sering mengacaukan pelatihan model RNN [24]. Dalam artikel ini, *F1-score* diperlakukan sebagai metrik utama saat membaca *cross-validation* karena klasifikasi

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

bersifat multikelas dan distribusi sentimen tidak seimbang [25]. Akurasi, presisi, *recall*, dan AUC tetap dipertahankan sebagai metrik pelengkap agar kemampuan diskriminasi model dapat dibaca secara lebih komprehensif.

2.7 Pengujian Model

Dua arsitektur diuji, yaitu LSTM dan GRU. Evaluasi akhir dilakukan pada data uji menggunakan akurasi, presisi, *recall*, *F1-score*, dan AUC [26]. Selain nilai metrik, dilakukan pengakjian terhadap *confusion matrix* mengenai resiko misklasifikasi sentimen.

2.8 Analisis Linguistik, Temporal, Spasial

Analisis linguistik dilakukan setelah data memperoleh label sentimen. Analisis lingusitik dilakukan untuk mengenali penyebab sentimen pada tiap *tweet* sehingga yang dihasilkan bukan sekedar sentimen terbanyak, tetapi isu pada tiap sentimennya serta aspek yang paling memengaruhinya seperti tokoh, tindakan, objek, dan hal lainnya. Analisis linguistik ini dilakukan untuk mengkaji setiap *tweet* secara utuh agar didapatkan kesimpulan yang lebih baik. Pertama, dilakukan n-gram digunakan untuk menemukan kata dan pasangan kata yang paling sering muncul pada setiap kelas sentimen. Kedua, POS *tagging* dipakai untuk membaca dominasi kelas kata seperti nomina, verba, dan adjektiva. Ketiga, *dependency parsing* digunakan untuk membentuk pola subjek–predikat–objek sehingga hubungan subjek dengan tindakan maupun objeknya dapat diidentifikasi [11]. Analisis temporal dilakukan dengan mengelompokkan *tweet* per bulan untuk melihat perubahan fokus percakapan dari Januari sampai Agustus 2025. Analisis spasial dilakukan pada tweet yang memuat keterangan lokasi dan kemudian dipetakan ke pulau besar di Indonesia. Kombinasi analisis ini dipakai untuk menghubungkan performa klasifikasi dengan konteks isu yang berkembang di masing-masing periode dan wilayah.

3. HASIL DAN PEMBAHASAN

3.1 Eksplorasi *Balancing*

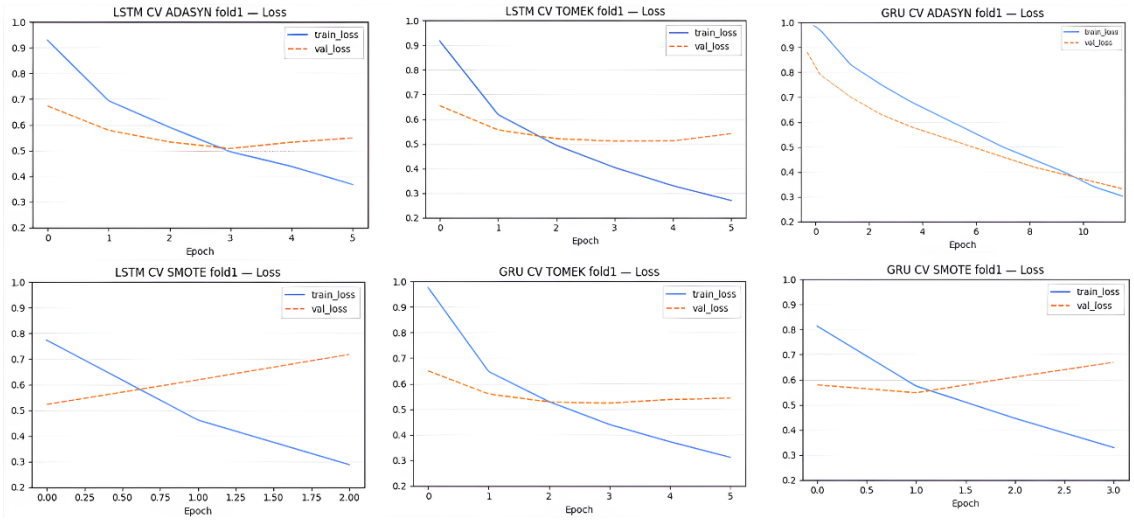
Pada tahapan *balancing data* dilakukan eksplorasi dnegan mengevaluasi beberapa skenario, yaitu *oversampling* dan *undersampling*. Pemilihan skenario final tidak hanya didasarkan pada nilai *F1-score* tertinggi, tetapi juga mempertimbangkan stabilitas pembelajaran dan indikasi *overfitting* yang diamati dari *learning curves*.

Tabel 6 Hasil Eksplorasi Teknik Balancing

LSTM				GRU			
Skenario	*Accuracy*	*F1-score*	Indikasi *Overfit*	Skenario	*Accuracy*	*F1-score*	Indikasi *Overfit*
Tomek	0.79729	0.793554	Ringan	Tomek	0.782465	0.77679	Ringan
ADASYN	0.78597	0.781187	Ringan	ADASYN	0.766232	0.76101	Tidak ada
SMOTE	0.77735	0.771669	Tinggi	SMOTE	0.776353	0.77475	Tinggi

Skenario yang dipilih adalah skenario yang menunjukkan performa validasi yang stabil serta memiliki selisih train dan validation yang lebih kecil. Pada Gambar 3, *balancing* dengan ADASYN menunjukkan gap paling kecil antara *train* dan *val*. balancing menggunakan ADASYN juga menghasilkan kurva terbaik. Hal tersebut tunjukan dengan menurunnya *train_loss* dan diikuti dengan penurunan *val_loss*. Oleh karena itu, penelitian ini difokuskan menggunakan ADASYN untuk *balancing* data.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255



Gambar 3. Hasil *Learning Curves* Eksplorasi *Balancing*

3.2 Kinerja Pelatihan dan Perbandingan Model

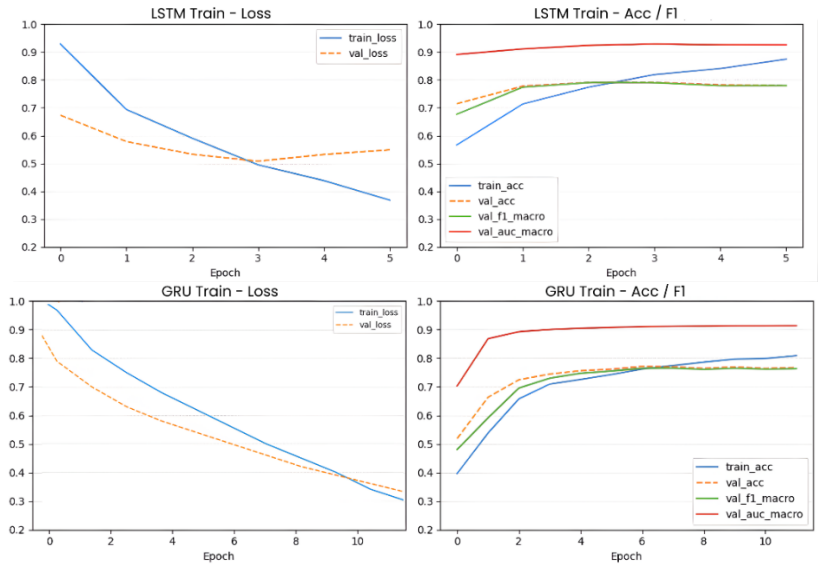
Pada tahapan *train*, model LSTM memiliki rata-rata akurasi 78,60%, presisi 78,28%, *recall* 78,33%, dan *F1-score* 78,12%, dengan performa terbaik pada *fold* ke-4 yang mencapai akurasi 80,16% dan *F1-score* 79,97%. GRU mencapai rerata akurasi 76,62%, presisi 76,07%, *recall* 76,31%, dan *F1-score* 76,10%. Selisih ini menunjukkan bahwa LSTM lebih unggul secara rerata, sedangkan GRU tetap kompetitif dan relatif stabil di berbagai pembagian *fold.* Pada Tabel 7 menunjukan nilai akurasi dan *f1-score* di tiap *fold* yang memperjelas selisih di tiap fold yang tidak jauh berbeda.

Tabel 7 Hasil Training LSTM dan GRU

LSTM			GRU		
Fold	*Accuracy*	*F1-score*	*Fold*	*Accuracy*	*F1-score*
1	0.791583166	0.789973	1	0.769539078	0.7652499
2	0.77254509	0.7700746	2	0.758517034	0.7505862
3	0.76753507	0.7581499	3	0.748496994	0.7439115
4	0.801603206	0.799731	4	0.766533066	0.7634915
5	0.774549098	0.7693644	5	0.771543086	0.7681763
6	0.791583166	0.778601	6	0.755511022	0.7417768
7	0.788577154	0.7821257	7	0.75751503	0.752992
8	0.793587174	0.7959309	8	0.77755511	0.7757484
9	0.785571142	0.7785441	9	0.77755511	0.7743522
10	0.79258517	0.7893761	10	0.779559118	0.7738753

Learning curve pada Gambar 4 memperlihatkan karakter pembelajaran yang sedikit berbeda. Pada LSTM, *validation loss* mulai cenderung meningkat setelah sekitar *epoch* ketiga meskipun *training loss* terus turun, sehingga model menunjukkan *overfitting* ringan yang masih dapat dikendalikan oleh *dropout* dan regularisasi L2. Sebaliknya, GRU menampilkan *validation loss* yang lebih stabil hingga sekitar *epoch* ke-6. Temuan ini menjelaskan mengapa GRU terlihat lebih rapi dari sisi proses belajar, tetapi LSTM tetap menghasilkan batas keputusan yang lebih baik.

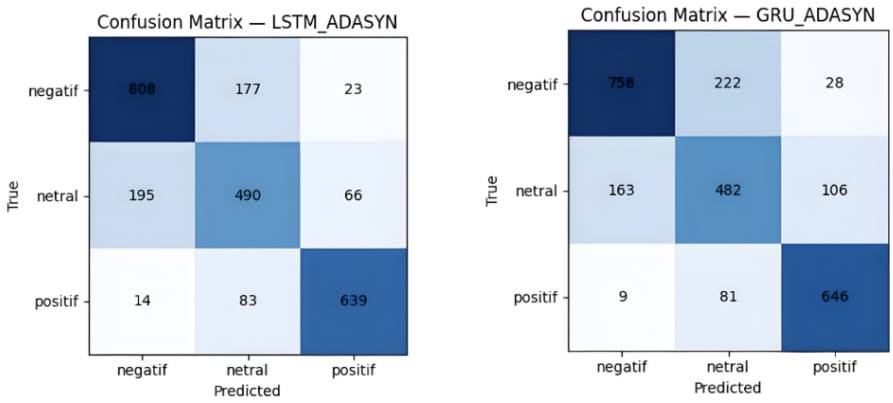
e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255



Gambar 4. Kurva Train LSTM dan GRU

3.3 Pengujian Model

Struktur kesalahan model dapat dilihat dengan *Confusion Matrix* pada Gambar 5. *Confusion matrix* merupakan tabel yang menunjukkan perbandingan antara informasi kelas aktual dan kelas prediksi dalam sebuah sistem klasifikasi. Tabel tersebut digunakan untuk menghitung kinerja proses klasifikasi atau model dengan nilai benar-salah yang telah diketahui.



Gambar 5. *Confusion Matrix*

Dari *confusion matrix*, LSTM menghasilkan prediksi benar sebanyak 808 *tweet* negatif, 490 netral, dan 639 positif. Kesalahan paling dominan terjadi pada perpindahan netral ke negatif dan negatif ke netral, sedangkan pergeseran negatif ke positif relatif kecil. Pada GRU, prediksi benar tercatat 758 *tweet* negatif, 482 netral, dan 646 positif serta kelas negatif masih cukup sering bergeser ke netral. Pola ini menegaskan bahwa tantangan utama bukan membedakan opini yang sangat mendukung dari yang sangat menolak, melainkan memisahkan kritik lunak dari pelaporan informatif. Hasil evaluasi pada Tabel 8 menunjukkan bahwa LSTM memberikan kinerja terbaik pada seluruh metrik utama. Keunggulan paling terlihat pada akurasi dan F1-score, sedangkan selisih *area under curve* (AUC) antara kedua model relatif kecil. Temuan ini sejalan dengan rerata validasi silang yang juga menempatkan LSTM di atas GRU, yakni sekitar 78,12% berbanding 76,10% pada *F1-score*. Dengan demikian, LSTM lebih layak dijadikan model utama untuk pembacaan sentimen lanjutan pada isu ini

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

Tabel 8 Perbandingan Kinerja Model pada Data Uji

Model	Akurasi	Presisi	*Recall*	*F1-score*	AUC
Long Short-Term Memory	77,63%	77,51%	77,40%	77,46%	91,54%
Gated Recurrent Unit	75,59%	75,24%	75,71%	75,40%	90,82%

3.4 Temuan Analisis Linguistik

Untuk mengkaji *tweet* lebih utuh, dilakukan analisis linguistik dengan N-gram, POS *Tagging*, *Depedency Parsing* untuk mengidentifikasi pola percakapan dan penyebab dari tiap sentimennya. Pada proses analisis linguistik digunakan data yang telah diberi label sebelumnya. Secara umum, percakapan didominasi sentimen negatif yang banyak menyoroti isu kualitas makanan dan anggaran, sementara sentimen positif muncul pada narasi manfaat gizi dan pemerataan program, dan sentimen netral cenderung berupa informasi pelaksanaan tanpa penilaian eksplisit.

3.4.1 Analisis N-gram

N-gram unigram dan bigram dianalisis untuk memudahkan identifikasi pola kata umum dalam setiap kategori guna melihat isu dari tiap sentimen. Berdasarkan hasil unigram dan bigram pada Tabel 9 yang menunjukkan bahwa sentimen negatif didominasi oleh kosakata yang berhubungan dengan kualitas makanan, dan tata kelola anggaran. Pada kelas netral, kata-kata yang dominan cenderung bersifat institusional dan informatif, misalnya "bgn", "daerah", "dapur", dan "kegiatan". Sebaliknya, sentimen positif diisukan memberi gambaran mengani narasi politik suatu tokoh, dan manfaat jangka panjang program yang ditunjukan dengan kata seperti "presiden prabowo", "masa depan", "pertumbuhan ekonomi", dan "generasi emas”.

Tabel 9 Hasil N-gram

Sentimen Negatif			
Unigram	Freq	Bigram	Freq
keracunan	419	gara gara	53
makanan	385	ulat sagu	50
anggaran	332	makanan basi	49
basi	306	bagi bagi	32
korupsi	226	efisiensi anggaran	30
Sentimen Netral			
Unigram	Freq	Bigram	Freq
bgn	138	bgn dadan	28
daerah	122	jawa timur	50
dapur	108	pemberian makan	26
kepala	81	ketua umum	26
kegiatan	74	sigit prabowo	26
Sentimen Positif			
Unigram	Freq	Bigram	Freq
prabowo	857	presiden prabowo	717
presiden	814	masa depan	126
papua	697	pertumbuhan ekonomi	105
dukung	478	siswa papua	93
cerdas	279	generasi emas	47

3.4.2 Analisis POS *Tagging*

Part of Speech (POS) *Tagging* merupakan teknik NLP dalam penandaan kelas kata yang diimplementasikan menggunakan *library* Stanza Bahasa Indonesia. Hasil POS *tagging* memperlihatkan bahwa nomina merupakan kelas kata yang paling dominan pada seluruh label,

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

diikuti verba. Dominasi yang ditunjukkan Tabel 10 menunjukkan bahwa percakapan publik dibentuk oleh pembicaraan mengenai aktor, objek, dan tindakan kebijakan. Adjektiva muncul lebih kuat pada *tweet* negatif dan positif daripada *tweet* netral, sehingga penilaian emosional terhadap program lebih banyak dibangun melalui kata sifat.

Tabel 10 Hasil POS *Tagging*

Label	Kata sifat dominan	Nomina dominan	Verba dominan	Makna yang dibentuk
Negatif	miskin (45), basi (39), aneh (27), bodoh (25), beda (19)	racun (138), duit (75), serangga (66), susu (58), korupsi (57)	pakai (81), bayar (54), racun (44), bawa (38), suruh (36)	Nada evaluatif dan menuduh, fokus pada mutu menu, anggaran, dan rasa tidak percaya.
Netral	nasional (23), lokal (23), awal (22), standar (21), umum (19)	menu (145), siswa (111), daerah (69), dapur (67), kepala (54)	tinjau (34), buka (29), naik (28), belajar (28), baca (25)	Bahasa administratif dan pelaporan, menekankan lokasi, pelaksana, dan kegiatan.
Positif	sehat (184), cerdas (66), sukses (40), cerah (38), bagus (35)	presiden (368), gizi (341), ekonomi (211), generasi (141), stunting (73)	lanjut (244), dukung (227), tumbuh (78), maju (71), bangun (43)	Menautkan program dengan masa depan anak dan pembangunan.

3.4.3 Analisis *Depedency Parsing*

Berdasarkan hasil analisis, informasi digabungkan dengan label sentimen untuk menghitung subjek, predikat, dan objek yang paling sering muncul dalam setiap kategori sentimen. Hal ini akan memungkinkan penjelasan yang lebih tepat tentang hubungan subjek, tindakan, dan objek yang paling sering muncul dalam *tweet* negatif, netral, dan positif. *Dependency parsing* pada Tabel 11 menegaskan bahwa pembeda utama sentimen bukan sekadar siapa subjeknya, melainkan kombinasi antara predikat dan objek. Sentimen negatif muncul ketika predikat diarahkan pada objek yang bernuansa risiko. Sebaliknya, sentimen positif terbentuk ketika verba terhubung dengan objek manfaat. Pada kelas netral, pola yang dominan berupa pelaporan aktivitas, seperti peninjauan sekolah, pembagian menu, atau pelaksanaan program di daerah tertentu.

Tabel 11 Hasil *Depedency Parsing* Level Kata

Label	Subjek dominan	Predikat dominan	Objek dominan
Negatif	anak (193), mbg (193), program (173), orang (112), pemerintah (105)	makan (739), cegah (694), bikin (236), pakai (213), buat (211)	mbg (691), program (583), rakyat (336), uang (291), menu (279)
Netral	program (122), mbg (89), anak (82), ketua (48), kita (43)	ada (427), datang (365), tanya (315), tinjau (265), jalan (76)	anak (488), mbg (382), sekolah (253), siswa (145), presiden (144)
Positif	program (510), presiden (209), mbg (174), dukung (60), kita (39)	lanjut (409), dukung (387), tingkat (213), tumbuh (136), beri (135)	anak (937), mbg (924), gizi (845), stunting (381), ekonomi (338)

Kombinasi tersebut menunjukkan bahwa percakapan negatif tidak hanya menyalahkan aktor, tetapi menyoroti apa yang dilakukan program terhadap objek yang dipersepsikan rentan atau dirugikan. Sebaliknya, pada kelas positif menegaskan bahwa objek lebih diagnostik dibanding subjek. Temuan Tabel 12 membantu menjelaskan mengapa salah klasifikasi terbesar terjadi pada

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

batas negatif dan netral, karena keduanya sering berbagi subjek, tetapi berbeda halus pada nada predikat dan objek.

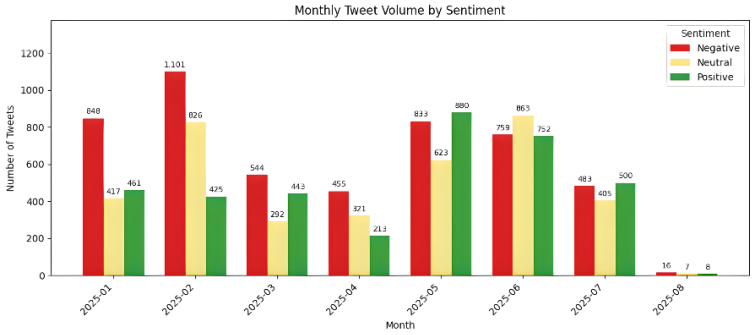
Tabel 12 Hasil *Depedency Parsing* Level Frasa

Label	Subjek dominan	Predikat dominan	Objek dominan
Negatif	Program mbg (299), orang tidak (127), orang miskin (107), pajak (106), ratus siswa sma jatinangor tolak mbg (80)	makan (739), cegah (694), bikin (236), pakai (213), buat (211)	mbg untuk anak (1524), program mbg (397), anggaran makan bergizi gratis (156), nasi bau (148), belalang gratis (279)
Netral	Ketum bhayangkari (75), sekolah pelaksana program makan bergizi gratis (69), wapres gibran rakabuming (60), nyonya juliati sigit prabowo (33)	ada (427), datang (365), tanya (315), tinjau (265), jalan (76)	makan bergizi gratis (260), program makan bergizi gratis (235), mbg untuk anak (214), pelaksana program makan (78), nasi bergizi gratis di sekolah (68)
Positif	presiden (301), program makan bergizi gratis (242), pemerintah (103), mbg untuk anak (101), yang menempatkan masyarakat (39)	lanjut (409), dukung (387), tingkat (213), tumbuh (136), beri (135)	program makan bergizi gratis (390), mbg untuk anak (257), masa depan anak (175), anak indonesia sehat cerdas (143), sejahtera anak di seluruh indonesia (100)

Frasa negatif memperjelas keluhan tentang kualitas makanan dan penggunaan dana publik. Frasa netral didominasi konstruksi yang jelas bersifat pemberitaan. Sementara itu, frasa positif membingkai MBG sebagai intervensi pembangunan, bukan sekadar distribusi makanan. Hal tersebut memperlihatkan bahwa frasa negatif cenderung berpusat pada insiden, frasa netral pada dokumentasi kegiatan, dan frasa positif pada narasi manfaat.

3.5 Dinamika Temporal per Bulan

Analisis temporal menunjukkan bahwa percakapan tentang Program Makan Bergizi Gratis sangat responsif terhadap peristiwa yang menyertainya yang ditunjukan pada Gambar 6.



Gambar 6. Histogram Dinamika Temporal

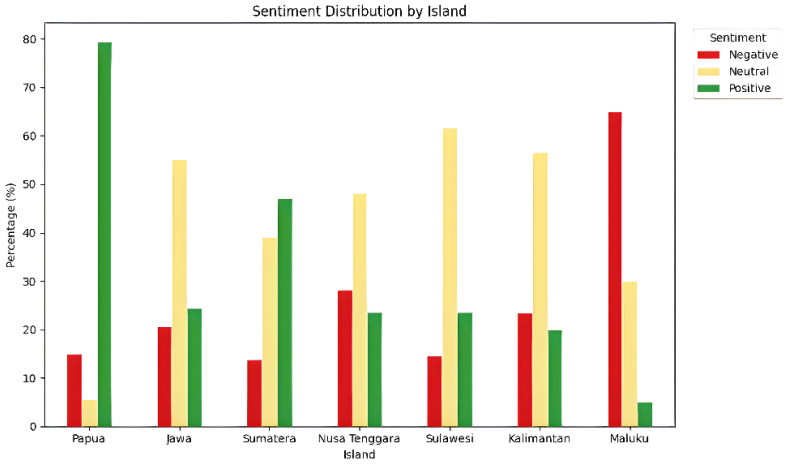
Volume *tweet* tertinggi terjadi pada Februari 2025, sedangkan pergeseran sentimen berjalan dinamis sepanjang periode penelitian. Secara umum, awal implementasi program berasosiasi dengan dominasi sentimen negatif karena isu kualitas makanan dan kritik terhadap kesiapan pelaksanaan lebih menonjol. Jika dibaca per fase, Januari–Februari dapat disebut sebagai fase *problem recognition*, ketika kasus keamanan makanan dan kualitas menu menjadi pendorong utama sentimen negatif. Februari merupakan bulan dengan volume *tweet* tertinggi, sehingga isu keracunan dan kesiapan pelaksanaan terlihat sangat dominan pada periode awal. Maret–April menunjukkan fase transisi karena percakapan mulai diisi oleh dua arus sekaligus yaitu kritik terhadap implementasi dan laporan kegiatan di sekolah atau daerah. Pada Mei–Juni, pembicaraan mengenai anggaran, efisiensi, dan prioritas belanja publik kembali menguat sehingga sentimen negatif meningkat. Juli memperlihatkan perubahan *framing* ketika kosakata

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

aspiratif seperti generasi emas, masa depan, dan sehatkan anak bangsa muncul lebih intens, sedangkan Agustus cenderung lebih tersebar dan netral karena percakapan banyak berisi pelaporan pelaksanaan serta evaluasi.

3.6 Perbedaan Spasial Antarwilayah

Analisis spasial dilakukan pada 1.089 *tweet* yang memuat keterangan lokasi sehingga hasilnya perlu dibaca sebagai kecenderungan, bukan representasi penuh seluruh wilayah yang sesuai ditunjukan pada Gambar 7.



Gambar 7. Histogram Tiap Pulau

Papua menampilkan kecenderungan sentimen positif paling kuat. Temuan ini selaras dengan narasi pemerataan manfaat program ke wilayah yang selama ini menghadapi keterbatasan akses pangan layak. Jawa memperlihatkan pola yang paling dinamis karena menjadi pusat pertemuan antara laporan pelaksanaan, dukungan politik, dan kritik kebijakan. Kalimantan dan Nusa Tenggara lebih sering menampilkan tweet netral yang bersifat informatif, sedangkan Maluku cenderung lebih kritis. Sumatera dan Sulawesi menunjukkan kecenderungan yang lebih positif pada beberapa periode, terutama ketika program dibingkai sebagai investasi sosial dan ekonomi. Perbedaan antarpulau menunjukkan bahwa konteks lokal memengaruhi cara publik menafsirkan program. Papua menjadi wilayah dengan sinyal positif paling kuat karena MBG lebih sering dibaca sebagai jawaban atas keterbatasan akses pangan layak. Hal ini konsisten dengan kemunculan n-gram "siswa papua" pada kelas positif.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa arsitektur *Long Short-Term Memory* (LSTM) berkinerja lebih baik daripada *Gated Recurrent Unit* (GRU) dalam memproses opini publik tentang program "Makan Bergizi Gratis" di media sosial X. Meskipun perbedaan kinerja tidak terlalu signifikan secara statistik, model LSTM memberikan nilai yang lebih stabil untuk akurasi, skor F1, dan AUC. Dari segi isi, keluhan publik seringkali berkaitan dengan kekhawatiran tentang kualitas makanan dan transparansi dalam pengelolaan anggaran. Di sisi lain, sentimen positif muncul dari narasi pemerataan akses terhadap makanan dan masa depan yang lebih baik bagi anak-anak di wilayah Papua. Analisis linguistik juga mengungkapkan fakta bahwa opini publik lebih kuat dipengaruhi oleh interaksi antara tindakan dan tujuan daripada hanya berfokus pada subjek yang melakukan tindakan tersebut. Perspektif publik ini tampak mudah berubah dan terus bergeser sebagai respon terhadap isu-isu lokal. Dari perspektif spasial, penduduk Papua menunjukkan tingkat minat dan dukungan tertinggi dibandingkan dengan wilayah lain. Temuan ini menunjukan bahwa evaluasi program pemerintah tidak hanya harus menghitung jumlah suara

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

yang mendukung dan menentang, tetapi juga mempertimbangkan konteks linguistik, waktu kejadian, dan lokasi asal opini tersebut.

Secara metodologis, pendekatan penelitian ini menawarkan perspektif yang jauh lebih mendalam. Kombinasi klasifikasi berbantuan komputer, analisis linguistik, dan pencatatan waktu dan lokasi membuat hasil evaluasi lebih bermakna daripada jika hanya skor akurasi algoritma yang disajikan. Pendekatan berlapis-lapis seperti ini sangat cocok untuk diterapkan dalam evaluasi kebijakan publik lainnya, karena dapat mengubah hasil yang dihasilkan mesin menjadi pesan otentik dari masyarakat. Oleh karena itu, hasil penelitian ini harus dipahami sebagai cerminan dinamika opini publik pada saat itu, dan bukan sebagai penilaian absolut atas keberhasilan program secara keseluruhan.

REFERENSI

[1] A. A. Merlinda and Y. Yusuf, "Analisis Program Makan Gratis Prabowo Subianto Terhadap Strategi Peningkatan Motivasi Belajar Siswa di Sekolah Tinjauan dari Perspektif Sosiologi Pendidikan," *Ranah Res. J. Multidiscip. Res. Dev.*, vol. 7, no. 2, pp. 1364–1373, 2025, doi: 10.38035/rrj.v7i2.1360.

[2] S. Rehani, "Social Media Personalization Algorithms and the Emergence of Filter Bubbles," *iSChannel*, vol. 15, no. 1, pp. 20–24, 2020.

[3] M. Cinelli, G. de Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, "The echo chamber effect on social media," in *Proceedings of the National Academy of Sciences of the United States of America*, 2021, pp. 1–8. doi: 10.1073/pnas.2023301118.

[4] A. Zhang, Z. Lipton, M. Li, and A. J. Smola, *Dive Into Deep Learning*, vol. 17, no. 5. 2023. doi: 10.1016/j.jacr.2020.02.005.

[5] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[6] A. Edward, B. J. Zevini, and H. M. Abiola, "Analyzing Election Sentiments in Tweets with Gated Recurrent Units," *Asian J. Res. Comput. Sci.*, vol. 16, no. 4, pp. 125–132, 2023, doi: 10.9734/ajrcos/2023/v16i4376.

[7] S. C. Amaria and N. Hidayati, "ANALISIS PROGRAM MAKAN BERGIZI GRATIS DENGAN SUPPORT VECTOR MACHINE (SVM) PADA APLIKASI X," *J. Sist. Inf.*, vol. 12, no. 2, pp. 16–24, 2025.

[8] P. Subarkah, A. N. Ikhsan, E. Anggraeni, and A. P. Sabaniyah, "Sentiment Perspective of Government's Free Nutritious Meal Policy on Social Media X using Indo-BERT and Bi-LTSM," *J. Technol. Informatics*, vol. 7, no. 2, pp. 201–210, 2025, doi: 10.37802/joti.v7i2.1065.

[9] Krisnawan, Z. A. Rabbani, Trimono, and M. Idhom, "IT-EXPLORE Analisis sentimen program makan bergizi gratis menggunakan bidirectional gated recurrent unit," *J. Penerapan Teknol. Inf. dan Komun.*, vol. 4, no. 3, pp. 282–294, 2025, doi: 10.24246/itexplore.v4i3.2025.pp282-294.

[10] N. G. Ramadhani and A. Khoirunnisa, "An Integrated Random Forest for Analyzing Public Sentiment on the 'Makan Bergizi Gratis' Program Nur," *J. Inf. Syst. Informatics*, vol. 7, no. 3, pp. 2314–2328, 2025, doi: 10.51519/journalisi.v6i2.759.

[11] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 2021, pp. 843–857. doi: 10.18653/v1/2020.aacl-main.85.

[12] H. Satria, *Cara Mendapatkan Data (Crawl) Twitter X - 30 Maret 2024*, (2024). [Online]. Available: https://youtu.be/BxryBUtUqEk?si=Lya8TrjSUKqwb3-f

[13] M. O. Ibrohim and I. Budi, "A Dataset and Preliminaries Study for Abusive Language Detection in Indonesian Social Media," in *Procedia Computer Science*, 2018, pp. 222–229. doi: 10.1016/j.procs.2018.08.169.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255

[14] N. A. Salsabila, Y. Ardhito Winatmoko, A. Akbar Septiandri, and A. Jamal, "Colloquial Indonesian Lexicon," in *Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018*, 2018, pp. 226–229. doi: 10.1109/IALP.2018.8629151.

[15] A. Abidin, "Kamus Custom Preprocessing Analisis Sentimen Indonesia." [Online]. Available: https://github.com/ArhamZaldin/Kamus-Custom-Preprocessing-Analisis-Sentimen-Indonesia/blob/main/sentiment-words.json

[16] D. H. Wahid and A. SN, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 10, no. 2, p. 207, 2016, doi: 10.22146/ijccs.16625.

[17] P. Ayuningtyas, S. Khomsah, and S. Sudianto, "Pelabelan Sentimen Berbasis Semi-Supervised Learning menggunakan Algoritma LSTM dan GRU," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 3, pp. 217–229, 2024, doi: 10.14421/jiska.2024.9.3.217-229.

[18] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychol. Bull.*, vol. 76, no. 5, pp. 378–382, 1971, doi: 10.1037/h0031619.

[19] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data Published by : International Biometric Society Stable URL : http://www.jstor.org/stable/2529310," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

[20] S. Szeghalmy and A. Fazekas, "A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning," *Sensors*, vol. 23, no. 4, pp. 1–27, 2023, doi: 10.3390/s23042333.

[21] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proceedings of the International Joint Conference on Neural Networks*, 2008, pp. 1322–1328. doi: 10.1109/IJCNN.2008.4633969.

[22] O. G. Al-Salih, D. Guangjian, W. J. Al-Mudhafar, and D. A. Wood, "Using extreme gradient boosting with Optuna hyperparameter tuning for efficient lost circulation prediction," *Energy Geosci.*, vol. 7, no. 2, p. 100540, 2026, doi: 10.1016/j.engeos.2026.100540.

[23] R. Belaroussi, S. C. Noufe, and F. Dupin, "Polarity of Yelp Reviews : A BERT – LSTM Comparative Study," *Big data Cogn. Comput.*, vol. 9, no. 5, pp. 1–25, 2025, doi: https://doi.org/10.3390/bdcc9050140.

[24] M. Kastrati, Z. Kastrati, A. Shariq, and I. Marenglen, "Leveraging distant supervision and deep learning for twitter sentiment and emotion classification," *J. Intell. Inf. Syst.*, vol. 62, no. 10, pp. 1045–1070, 2024, doi: 10.1007/s10844-024-00845-0.

[25] D. Sartika and I. Saluza, "Penerapan Metode Principal Component Analysis (PCA) Pada Klasifikasi Status Kredit Nasabah Bank Sumsel Babel Cabang KM 12 Palembang Menggunakan Metode Decision Tree," *Generic J.*, vol. 14, no. 2, pp. 45–49, 2022, doi: 10.18495/generic.v14i2.130.

[26] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *Bio Data Min.*, vol. 16, no. 4, pp. 1–23, 2023, doi: 10.1186/s13040-023-00322-4.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16255