



Analisis Sentimen Program Makan Bergizi Gratis (MBG) untuk Anak Sekolah Menggunakan Algoritma *Support Vector Machine*

Muhammad Rizqi Warsita¹, Chaerur Rozikin², Pandapotan Kristian Sitorus³

^{1,2,3} Universitas Singaperbangsa Karawang

Received: 10 September 2026
Revised: 15 September 2026
Accepted: 01 Oktober 2026

Abstract

The Support Vector Machine with a linear kernel has proven to be highly effective and balanced in classifying public perceptions of the Free Nutritious Meals Programme (MBG). This conclusion was drawn from an analysis of public opinion on the social media platform X, which aimed to map public sentiment towards the programme. The analysis was conducted using a linear kernel SVM algorithm with the SEMMA methodology, utilising 1,000 initial data points from X, which were filtered down to 578 binary data points (401 positive and 177 negative) through a TF-IDF-based cleaning and labelling process. The data was split in an 80:20 ratio and tested across five variations of the hyperparameter C. The best configuration was achieved at C=1, with a global accuracy of 96.55 per cent, a negative F1-score of 94.44 per cent, and a positive F1-score of 97.50 per cent. In terms of content, positive sentiment centred on improvements in children's nutrition (keywords 'nutrition' and 'food'), whilst negative sentiment related to the risk of food poisoning and budget efficiency (keywords 'poison' and 'budget'). Given this high and balanced performance across both positive and negative classes, the linear kernel SVM is considered reliable for the task of classifying public policy perceptions.

Keywords: Sentiment Analysis, Free Nutritious Meals, Support Vector Machine, Linear Kernel, SEMMA.

(*) Corresponding Author:

rizqyyahoo1@gmail.com¹, chaerur.rozikin@staff.unsika.ac.id²,
pandachristian27@gmail.com³

How to Cite: Rizqi Warsita, M., Rozikin, C., & Kristian Sitorus, P. (2026). Analisis Sentimen Program Makan Bergizi Gratis (MBG) untuk Anak Sekolah Menggunakan Algoritma Support Vector Machine. *Jurnal Ilmiah Wahana Pendidikan*, 12(10.D), 197-206. Retrieved from <https://jurnal.peneliti.net/index.php/JIWP/article/view/14626>

PENDAHULUAN

Program Makan Bergizi Gratis (MBG), yang dirancang sebagai strategi nasional untuk memperbaiki gizi anak sekolah, menekan angka *stunting*, dan meningkatkan kualitas SDM, justru memunculkan polarisasi opini di kalangan publik. Sebagian pihak mendukung program ini demi perbaikan gizi, sementara sebagian lain mengkritik kesiapan anggaran, jalur distribusi, dan efektivitas implementasinya di berbagai daerah. Perdebatan semacam ini banyak mengemuka di platform X, yang telah menjadi wadah krusial bagi publik untuk menyuarakan opini, kritik, dan tanggapan langsung terhadap kebijakan pemerintah. Peran X sebagai ruang diskursus publik ini tidak lepas dari lonjakan signifikan penggunaan media sosial di Indonesia, yang menurut laporan We Are Social mencapai 191 juta pengguna atau 68% populasi pada tahun 2025.

Support Vector Machine kernel linear terbukti efektif dan seimbang dalam mengklasifikasikan sentimen publik. Pada konfigurasi terbaik (C=1), model ini

mencapai akurasi global 96,55%, dengan F1-Score kelas negatif 94,44% dan kelas positif 97,50%. Capaian ini diperoleh melalui metodologi SEMMA, berbekal 1.000 data awal dari platform X yang disaring menjadi 578 data biner (401 positif, 177 negatif) lewat pembersihan dan pelabelan berbasis TF-IDF, lalu diuji dengan pembagian data 80:20 pada lima variasi nilai hyperparameter C. Secara isi, sentimen positif berpusat pada kata kunci "gizi" dan "makan" yang mencerminkan apresiasi terhadap perbaikan nutrisi anak, sedangkan sentimen negatif berpusat pada kata kunci "racun" dan "anggar" yang mencerminkan kekhawatiran soal keamanan pangan dan efisiensi anggaran.

Hasil ini menjawab tujuan penelitian, yaitu menganalisis sentimen publik di platform X terhadap Program Makan Bergizi Gratis (MBG) menggunakan algoritma SVM kernel linear berbasis metodologi SEMMA, sekaligus memberi manfaat berupa gambaran persepsi masyarakat sebagai bahan evaluasi kebijakan pemerintah dan referensi bagi riset analisis sentimen kebijakan publik di Indonesia. Penelitian ini relevan dilakukan karena, meski penerapan SVM pada isu kebijakan publik secara umum sudah banyak dikaji, penelitian khusus mengenai program MBG masih sangat terbatas.

Pemilihan kernel linear didasarkan pada keunggulannya saat dikombinasikan dengan pembobotan TF-IDF, yang menghasilkan model sederhana namun memiliki generalisasi baik. Algoritma SVM sendiri dipilih karena performanya yang unggul pada data teks berdimensi tinggi dan tugas klasifikasi biner menjadikannya solusi teknologi berbasis Analisis Sentimen (cabang dari *Natural Language Processing*) yang mampu mengelompokkan teks ke kategori positif, negatif, atau netral secara otomatis. Solusi otomatis ini dibutuhkan karena volume komentar di media sosial sangat masif, sehingga analisis opini secara manual menjadi tidak efektif dan memakan banyak waktu.

Volume komentar yang masif ini tidak lepas dari polarisasi opini publik terhadap Program MBG: sebagian kalangan mengapresiasi program demi perbaikan gizi, sementara sebagian lain mengkritik ketidaksiapan anggaran, rumitnya mekanisme distribusi, dan efektivitas implementasi yang berbeda-beda antardaerah. Program MBG sendiri dirancang sebagai strategi nasional bagi anak sekolah untuk menaikkan kualitas gizi, menekan angka stunting, dan mendongkrak kualitas SDM masa depan sebuah isu yang wajar memicu tanggapan beragam. Perdebatan ini banyak mengemuka di platform X, yang telah bergeser menjadi wadah krusial bagi publik untuk menyuarakan opini, kritik, dan tanggapan langsung terhadap kebijakan pemerintah. Peran X sebagai ruang diskursus ini didorong oleh lonjakan signifikan penggunaan media sosial di Indonesia, yang menurut laporan *We Are Social* mencapai 191 juta pengguna atau 68% populasi pada tahun 2025.

TINJAUAN PUSTAKA

A. RESEARCH GAP

Penelitian ini mengisi celah yang ada dengan mengintegrasikan metodologi SEMMA dan algoritma Support Vector Machine (SVM) kernel linear untuk memetakan sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG) di platform X. Integrasi ini diperlukan karena beberapa kesenjangan masih terbuka dalam literatur yang ada. Pertama, studi khusus mengenai sentimen publik terhadap kebijakan strategis MBG saat ini masih sangat terbatas. Kedua, banyak riset

terdahulu hanya berfokus pada hasil akhir klasifikasi tanpa memaparkan tahapan pengolahan data secara komprehensif menggunakan metodologi data mining tertentu. Selain itu, penggunaan metodologi SEMMA dalam analisis sentimen terkait kebijakan publik juga masih relatif jarang ditemukan. Kesenjangan-kesenjangan ini sejalan dengan pola pada penelitian analisis sentimen media sosial sebelumnya, yang mayoritas berfokus pada ulasan produk, layanan digital, dan isu politik nasional, dengan metode seperti Naïve Bayes, KNN, dan Decision Tree.

B. STATE OF THE ART

Penelitian ini memperbarui pemetaan ilmiah dengan menerapkan metodologi SEMMA dan algoritma SVM kernel linear pada analisis kebijakan publik Program Makan Bergizi Gratis (MBG), sebuah objek kajian yang masih langka diteliti. Pemilihan SVM didasarkan pada kajian terkini yang membuktikan algoritma ini tetap menjadi salah satu yang terbaik untuk klasifikasi teks karena sifatnya yang stabil dan mampu menangani data berdimensi tinggi. Lebih spesifik, penelitian terbaru menunjukkan bahwa perpaduan pembobotan TF-IDF dan kernel linear pada SVM menghasilkan akurasi yang lebih tinggi untuk data teks media sosial dibandingkan jenis kernel lainnya. Metodologi SEMMA sendiri dipilih karena menyediakan tahapan data mining yang sangat sistematis, mulai dari pengambilan data hingga proses evaluasi model.

C. ANALISIS SENTIMEN

Penelitian ini memperbarui pemetaan ilmiah dengan menerapkan SEMMA dan SVM kernel linear pada analisis kebijakan publik Program Makan Bergizi Gratis (MBG), objek yang masih langka diteliti. SVM sendiri adalah algoritma supervised learning untuk klasifikasi dan regresi yang bekerja mencari hyperplane optimal pemisah antar-kelas data, dengan tujuan memaksimalkan margin antar-kelas agar model punya generalisasi baik. Karakteristik ini menjadikannya efektif memproses data teks berdimensi tinggi dengan performa klasifikasi tetap tinggi sejalan dengan kajian terkini yang membuktikan SVM tetap salah satu algoritma terbaik untuk klasifikasi teks karena stabil dan tangguh pada dimensi tinggi. Pada data teks hasil pembobotan TF-IDF, sifatnya yang umumnya linearly separable membuat kernel linear mampu memberi hasil optimal dengan kompleksitas komputasi lebih rendah, dan penelitian terbaru menunjukkan kombinasi TF-IDF dengan kernel linear ini menghasilkan akurasi lebih tinggi untuk data teks media sosial dibanding kernel lain. Metodologi SEMMA dipilih karena menyediakan tahapan data mining sistematis dari pengambilan data hingga evaluasi model.

D. TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)

Pembobotan TF-IDF sangat populer digunakan dalam analisis sentimen karena mampu menghasilkan representasi fitur yang efektif dan akurat pada data teks. Efektivitas ini diperoleh dari cara kerjanya, yaitu menghitung frekuensi kemunculan kata dalam suatu dokumen (*Term Frequency*), lalu mengurangi bobot kata tersebut apabila kata itu terlalu sering muncul di seluruh dokumen (*Inverse Document Frequency*). Melalui mekanisme inilah TF-IDF berfungsi sebagai metode pembobotan kata untuk mengukur tingkat kepentingan suatu kata dalam sebuah dokumen terhadap keseluruhan kumpulan dokumen (dataset).

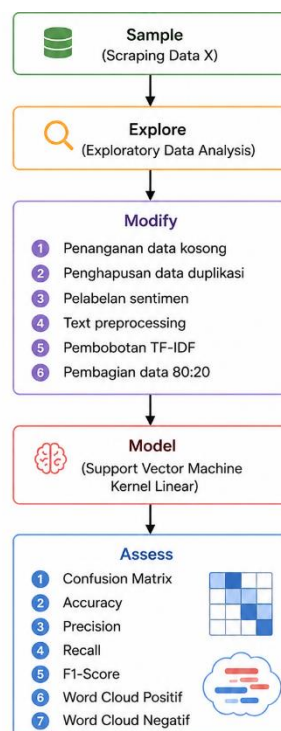
E. METODOLOGI SEMMA

SEMMA merupakan metodologi data mining yang dikembangkan oleh SAS Institute dan terdiri atas lima tahapan utama, yaitu Sample, Explore, Modify, Model, dan Assess.

- 1) Sample, yaitu proses pengumpulan data komentar dari platform X sebanyak 1.000 data.
- 2) Explore, yaitu melakukan Exploratory Data Analysis (EDA) untuk memahami karakteristik data.
- 3) Modify, meliputi penanganan data kosong dan duplikasi, pelabelan sentimen menggunakan kata kunci, preprocessing text, pembobotan TF-IDF, dan pembagian data latih dan data uji sebesar 80:20.
- 4) Model, yaitu pembangunan model klasifikasi menggunakan algoritma Support Vector Machine dengan kernel linear.
- 5) Assess, yaitu evaluasi model menggunakan confusion matrix dan visualisasi word cloud untuk sentimen positif dan negatif.

METODE PENELITIAN

Seluruh rangkaian alur penelitian ini disajikan dan dapat dilihat pada Gambar 1, yang menggambarkan tahapan sistematis mulai dari proses pengolahan data hingga evaluasi model klasifikasi sentimen. Tahapan sistematis inilah alasan dipilihnya metodologi SEMMA (*Sample, Explore, Modify, Model, Assess*), sebuah metodologi yang dikembangkan oleh SAS Institute, sebagai kerangka penelitian ini.



Gambar 1. Metodologi Penelitian

A. Sample

Pada tahap ini dilakukan pengambilan data komentar dari platform media sosial X mengenai Program Makan Bergizi Gratis (MBG) menggunakan teknik

scraping. Jumlah data yang digunakan dalam penelitian ini sebanyak 1.000 data komentar.

B. Explore

Karakteristik dataset, distribusi data, dan pola awal dari data ditemukan melalui tahap Exploratory Data Analysis (EDA) yang dilakukan pada dataset komentar yang telah dikumpulkan.

C. Modify

Tahap modify meliputi:

- a. Penanganan data kosong (missing values);
- b. Penghapusan data duplikasi;
- c. Pelabelan sentimen menggunakan pendekatan berbasis kata kunci;
- d. Text preprocessing yang terdiri dari cleaning text, case folding, tokenizing, stopword removal, dan stemming;
- e. Transformasi data menggunakan pembobotan Term Frequency-Inverse Document Frequency (TF-IDF);
- f. Pembagian data menjadi data latih dan data uji dengan rasio 80:20.

D. Model

Model klasifikasi sentimen dibangun berdasarkan data yang telah dipersiapkan, menggunakan algoritma Support Vector Machine (SVM) dengan jenis kernel linear.

F. Assess

Kata-kata yang paling dominan pada masing-masing kategori sentimen positif dan negatif dipetakan menggunakan *Word Cloud*. Pemetaan ini melengkapi evaluasi kinerja model klasifikasi, yang diukur menggunakan komponen *Confusion Matrix* melalui perhitungan nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

HASIL & PEMBAHASAN

Hasil dan pembahasan untuk Analisis Sentimen yang sudah dilakukan yakni sebagai berikut.

A. SAMPLE

Sebanyak 1.000 komentar opini masyarakat mengenai pelaksanaan Program Makan Bergizi Gratis (MBG) berhasil dikumpulkan dan berperan sebagai dataset awal untuk seluruh rangkaian proses analisis sentimen selanjutnya. Data tersebut diperoleh melalui proses *scraping* pada platform media sosial X (Twitter) menggunakan kata kunci spesifik yang berkaitan dengan program MBG.

B. EXPLORE

Dataset dinyatakan bersih dan valid untuk langsung dilanjutkan ke tahap pengolahan data berikutnya. Kesimpulan ini diperoleh dari pemeriksaan karakteristik awal dataset melalui Exploratory Data Analysis (EDA), yang menunjukkan dataset berdimensi 1.000 baris data dan 6 kolom tanpa ditemukan data kosong (missing value) maupun data duplikat pada kolom teks. Detail karakteristik awal dataset ini ditampilkan pada Gambar 2.

	Properti Dataset	Nilai Analisis
0	Jumlah Total Baris	1000
1	Jumlah Total Kolom	6
2	Data Kosong pada Teks	0
3	Data Duplikat pada Teks	0

Gambar 2. Hasil Explore Dataset

C. MODIFY

Rangkuman dari seluruh proses tahap Modify ini tercantum pada Gambar 2. Proses tersebut menghasilkan pembagian dataset menggunakan metode *train-test split* dengan teknik *stratified sampling* (proporsi 80:20), menghasilkan 462 data latih dan 116 data uji, setelah sebelumnya teks ditransformasikan ke bentuk numerik menggunakan metode TF-IDF dengan batas maksimum 1.500 fitur kata. Transformasi ini dilakukan pada 578 data valid terdiri dari 401 sentimen positif dan 177 sentimen negatif setelah data berlabel netral dieliminasi karena penelitian hanya berfokus pada sentimen positif dan negatif. Data tersebut sebelumnya melalui tahapan *text preprocessing* berupa *cleaning text*, *case folding*, *stopword removal*, dan *stemming* menggunakan pustaka Sastrawi, serta proses pembersihan dan pelabelan yang mencakup penanganan data kosong (*missing value*) dan duplikat pada kolom Tweet Text, disertai pelabelan sentimen otomatis dengan pendekatan berbasis kata kunci (*rule-based labeling*). Seluruh rangkaian ini bertujuan mempersiapkan data mentah agar siap diproses oleh model algoritma Support Vector Machine (SVM).

	Tahapan Hasil Pemrosesan	Jumlah Baris Data
0	Total Data Bersih setelah Pembersihan	578
1	Sentimen Positif (Kata Kunci)	401
2	Sentimen Negatif (Kata Kunci)	177
3	Data Latih (80%)	462
4	Data Uji (20%)	116

Gambar 3. Hasil Modify

D. MODEL

Nilai $C=0,5$ keluar sebagai parameter terbaik dengan perolehan akurasi mencapai 96,55%. Hasil ini diperoleh dari pengujian terhadap 5 variasi nilai hyperparameter C , yaitu 0,1; 0,5; 1,0; 5,0; dan 10,0, yang dilakukan sebagai bagian dari optimasi model klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM) beralaskan kernel linear. Detail performa dan perbandingan dari masing-masing nilai hyperparameter C tersebut dimuat pada Gambar 4.

	Nilai Parameter C	Akurasi Uji (%)
0	0.1	68.97%
1	0.5	96.55%
2	1.0	94.83%
3	5.0	94.83%
4	10.0	94.83%

Gambar 4. Hasil Pelatihan

Model telah berhasil diselesaikan dengan baik dan siap dilanjutkan ke tahap evaluasi berikutnya (Assess). Model final yang ditetapkan dalam penelitian ini adalah Support Vector Machine (SVM) dengan jenis kernel linear dan nilai hyperparameter $C=0,5$, dengan representasi data teks ke bentuk numerik menggunakan metode TF-IDF Vectorizer. Rincian mengenai arsitektur dan spesifikasi model akhir ini disajikan pada Gambar 5.

	Komponen Model Klasifikasi	Konfigurasi / Parameter
0	Algoritma Utama	Support Vector Machine (SVM)
1	Tipe Kernel	Linear Kernel
2	Nilai C Terpilih (Terbaik)	0.5 (Akurasi: 96.55%)
3	Metode Fitur	TF-IDF Vectorizer
4	Status Pelatihan	Sukses Terlatih & Siap

Gambar 5. Spesifikasi Model

E. ASSESS

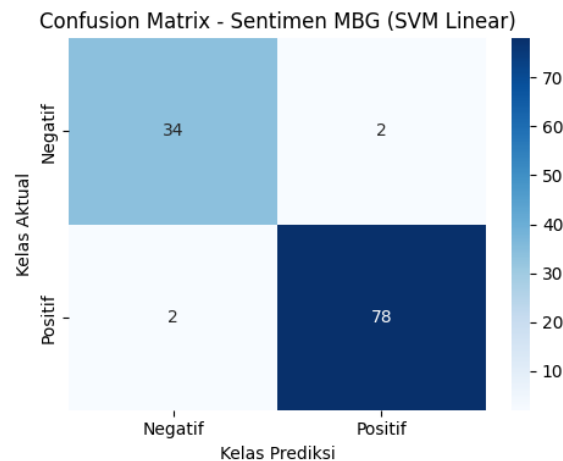
Seluruh rangkuman performa pengujian disajikan secara detail pada Gambar 6 dan Gambar 7. Performa tersebut diukur dari evaluasi klasifikasi model SVM kernel linear beserta analisis frekuensi kata guna memetakan tren opini publik, menggunakan 116 data uji yang terdiri dari 36 data sentimen negatif dan 80 data sentimen positif.

	Metrik Evaluasi Performa	Nilai Desimal	Nilai Persentase (%)
0	Accuracy (Akurasi Global)	0.9655	96.55%
1	Precision (Ketepatan Prediksi)	0.9750	97.50%
2	Recall (Sensitivitas Model)	0.9750	97.50%
3	F1-Score (Nilai Keseimbangan)	0.9750	97.50%

Gambar 6. Metrik Evaluasi

	precision	recall	f1-score	support
Negatif	0.9444	0.9444	0.9444	36.0000
Positif	0.9750	0.9750	0.9750	80.0000
accuracy	0.9655	0.9655	0.9655	0.9655
macro avg	0.9597	0.9597	0.9597	116.0000
weighted avg	0.9655	0.9655	0.9655	116.0000

Gambar 7. Metrik Klasifikasi



Gambar 8. Confussion Matrix

Ekstraksi kecenderungan kata kunci dilakukan guna membedah substansi opini masyarakat terhadap Program Makan Bergizi Gratis (MBG), dengan daftar 5 kata teratas yang paling mendominasi pada masing-masing kelompok sentimen (positif dan negatif) disajikan pada Gambar 9. Analisis ini melengkapi hasil evaluasi model, di mana nilai rata-rata tertimbang (weighted average) yang mencapai 96,55% menandakan model memiliki kemampuan generalisasi tinggi serta terhindar dari isu bias data. Kemampuan generalisasi ini tercermin dari kinerja model yang sangat stabil pada kedua kelas target, dengan F1-Score sebesar 94,44% untuk sentimen negatif dan 97,50% untuk sentimen positif, serta akurasi global yang tinggi, yaitu 96,55%.

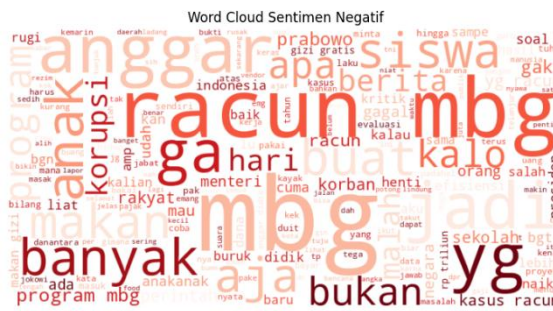
	Top Kata Positif	Frekuensi Positif	Top Kata Negatif	Frekuensi Negatif
0	gizi	380	mbg	200
1	mbg	366	racun	109
2	makan	300	yg	54
3	program	287	anggar	49
4	gratis	258	jadi	39

Gambar 9. Top 5 Kata teratas



Gambar 10. Word Cloud Sentimen Positif

Pada Gambar 10 dan Gambar 11 menampilkan word cloud persebaran kata positif dan negatif dan model yang telah dilatih.



KESIMPULAN

SARAN

REFERENCES

- Joachims, T. (1998). Text categorization with Support Vector Machines: Learning with many relevant features. *Proceedings of the European Conference on Machine Learning (ECML)*, 137–142.
- Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan Republik Indonesia. (2024). Program Makan Bergizi Gratis untuk Generasi Emas Indonesia.
- Koto, F., & Rahmaningtyas, G. (2017). InSet Lexicon: Evaluation of Indonesian sentiment analysis. *Proceedings of the International Conference on Asian Language Processing*, 391–394.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. California: Morgan & Claypool Publishers.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113.
- Moraes, R., Valiati, J., & Neto, W. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, 40(2), 621–633.
- Nasrullah, R. (2020). *Media Sosial: Perspektif Komunikasi, Budaya, dan Sosioteknologi*. Bandung: Simbiosis Rekatama Media.
- Pratama, A., & Wijaya, D. (2023). Sentiment analysis of public policies on social media using machine learning approaches. *Journal of Information Systems*, 19(2), 88–98.
- Putra, R. A., & Siregar, H. (2021). Analisis sentimen pada media sosial Twitter menggunakan metode Support Vector Machine. *Jurnal Teknologi dan Ilmu Komputer*, 18(1), 33–40.
- Ramos, J. (2003). Using TF-IDF to determine word relevance in document queries. *Proceedings of the First Instructional Conference on Machine Learning*, 29–48.
- SAS Institute. (2018). *SEMMA Methodology and Data Mining Process*.
- Singh, J., & Kumari, P. (2023). Sentiment analysis using TF-IDF and linear SVM for social media text classification. *International Journal of Intelligent Systems and Applications*, 15(2), 51–60.
- Suryani, D., & Prasetyo, A. (2024). Analisis persepsi masyarakat terhadap kebijakan makan bergizi gratis di Indonesia. *Jurnal Kebijakan Publik Indonesia*, 9(2), 112–121.
- Vapnik, V. N. (1998). *Statistical Learning Theory*. New York: Wiley.
- We Are Social. (2025). *Digital 2025: Indonesia*.