UNVEILING PUBLIC SENTIMENT: SUPPORT VECTOR MACHINE AND DECISION TREE ANALYSIS OF INSTAGRAM RESPONSES TO IMPLEMENTATION OF FREE NUTRITIOUS MEAL PROGRAM

Fitri Amelia[1*], Ayun Farichah Khoniaro[2], Rina Tisna Nurasih[3], Wilda Yulia Rusyida[4]
[1234]*Sains Data, Universitas Islam Negeri K.H Abdurrahman Wahid Pekalongan, Indonesia*
* ftramlia11@gmail.com

ABSTRAK

Pemerintah telah meluncurkan Program Makan Bergizi Gratis sebagai upaya strategis untuk menurunkan angka stunting dan meningkatkan kualitas gizi masyarakat. Namun, kebijakan ini memicu berbagai respons publik di media sosial, termasuk Instagram. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat dengan memanfaatkan algoritma *Support Vector Machine* dan *Decision Tree*. Data diperoleh dari 2.327 komentar bertagar #MakanBergiziGratis yang dikumpulkan melalui proses *scraping*, kemudian dianalisis melalui tahapan praproses teks, pembobotan TF-IDF, dan klasifikasi menggunakan *Python*. Sentimen dikategorikan menjadi positif dan negatif. Hasil analisis menunjukkan bahwa sentimen negatif sangat mendominasi (98,1%), dengan topik utama kekhawatiran terhadap pelaksanaan dan transparansi anggaran. Hasil menunjukkan kedua algoritma memiliki akurasi tinggi (99%). Pada sentimen positif, *Support Vector Machine* unggul dalam *precision* (1,00) tetapi memiliki *recall* rendah (0,44), sedangkan *Decision Tree* menunjukkan performa lebih seimbang dengan *precision* (0,83) dan *recall* (0,56). Temuan penelitian ini diharapkan dapat membantu pembuat kebijakan dalam mengevaluasi serta menyempurnakan implementasi program agar lebih efektif di masa mendatang.

Kata kunci: Analisis sentimen, *Decision Tree*, Instagram, Makan Bergizi Gratis, *Support Vector Machine*

1 PENDAHULUAN

Format Makan bergizi gratis menjadi salah satu program unggulan oleh pemerintah indonesia untuk meningkatkan kualitas gizi masyarakat. Program ini dirancang sebagai upaya strategis untuk mengurangi prevelensi angka stunting, meningkatkan mutu gizi, serta menunjang prestasi akademik siswa di berbagai tingkat pendidikan (Ritonga & Sazali, 2025). Berdasarkan hasil survei Status Gizi Indonesia (SSGI) tahun 2022, angka prevelensi stunting di Indonesia mencapai 21,6%. Hal ini menunjukkan bahwa prevalensi stunting di indonesia sudah mengalami penurunan dibandingkan dengan tahun sebelumnya yang mencapai 24,4% (Putriyekti et al., 2025). Namum penurunan angka prevelensi tersebut masih jauh dibawah standar *World Health Organization* (WHO) yang menetapkan bahwa prevelensi stunting harus berada pada ambang batas maksimal 20% (Pambudi, 2024). Prevelensi stunting yang tinggi di indonesia dapat membawa konsekuensi jangka panjang yang serius untuk kesehatan individu maupun dalam aspek pembangunan nasional. Anak-anak yang menderita stunting tidak hanya mengalami kesulitan dalam perkembangan fisik mereka, tetapi juga lebih rentan terhadap gangguan kognitif, yang dapat menyebabkan kualitas sumber daya manusia lebih rendah di masa depan (Merlinda & Yusuf, 2025).

Data dari Kementerian Kesehatan dan Kemenko Pembangunan Manusia dan Kebudayaan juga menunjukkan bahwa 41% siswa di indonesia mengalami kelaparan saat belajar di sekolah, sehingga hal tersebut berdampak negatif pada prestasi akademik siswa. Data dari Badan Pusat Statistik dan *World Health Organization* (2018) juga mengungkapkan bahwa sebagian besar masyarakat Indonesia masih mengalami kesulitan dalam memperoleh makanan bergizi, yang berdampak langsung pada kondisi kesehatan mereka. Program penyediaan makanan bergizi

gratis merupakan salah satu solusi penting untuk mengurangi ketimpangan ini dengan memperluas akses nutrisi yang cukup, khususnya bagi kelompok rentan. Menurut Roringkon dkk. (2022), program semacam ini dapat secara signifikan meningkatkan kualitas kesehatan masyarakat, sehingga menghasilkan sumber daya manusia (SDM) yang lebih sehat, produktif, dan mampu mendukung ketahanan nasional. Jika dilaksanakan dengan baik, inisiatif ini tidak hanya mengatasi masalah gizi buruk, tetapi juga menjadi landasan bagi terciptanya generasi mendatang yang lebih tangguh, baik dari segi kesehatan, pendidikan, maupun perannya dalam memperkuat pertahanan negara (Lestari et al., 2024). Target program ini tidak hanya untuk anak-anak sekolah, tetapi juga meliputi balita, ibu hamil, dan ibu menyusui (Merlinda & Yusuf, 2025).

Meskipun maksud dari program ini baik, akan tetapi program ini banyak menuai berbagai kontroversi pada media sosial, salah satunya di instagram, bahkan saat sebelum program makan bergizi gratis ini direalisasi. Media sosial memiliki peran penting dalam menyebarluaskan informasi terbaru terkait program-program pemerintah dan menjadi sebuah forum diskusi publik mengenai pendapat masyarakat terkait program tersebut. Respon dari masyarakat terhadap program makan bergizi gratis sangat beragam, khususnya saat program makan bergizi gratis ini resmi diluncurkan pada tanggal 6 Januari 2025 (Purwanti & Sugiyono, 2024). Dengan beragamnya respon masyarakat terhadap program ini, polarisasi opini publik semakin meningkat. Oleh karena itu, analisis sentimen diperlukan untuk mengetahui bagaimana masyarakat melihat kebijakan ini secara lebih objektif (Putriyekti et al., 2025). Analisis sentimen adalah upaya untuk menemukan dan mengategorisasi sudut pandang yang diungkapkan oleh seseorang dalam suatu grup teks, terutama untuk memastikan reaksi atau pendapat masyarakat terhadap program makan bergizi gratis ini, apakah positif, negatif, ataupun netral. Metode proses ini umumnya diterapkan pada berbagai aspek seperti pelayanan pelanggan maupun penelitian sosial untuk dapat memahami pandangan dan respons masyarakat mengenai produk atau layanan (Saleh & Imanda, 2025). Dalam penelitian ini akan digunakan analisis sentimen dengan metode algoritma SVM dan *Decision Tree*. SVM merupakan salah satu metode untuk mengklasifikasikan data opini, yang mana teknik ini tergolong baru dalam melakukan prediksi, baik dalam melakukan klasifikasi maupun regresi (Arrafiq & Kurniawan, 2024). SVM bekerja dengan cara memisahkan data menggunakan garis pembatas yang disebut hyperline, garis untuk membedakan kategori sentimen positif dan negatif berdasarkan pola yang dihasilkan dari proses *machine learning* (Purwanti & Sugiyono, 2024). Algoritma SVM adalah metode yang memiliki kinerja paling tingi dibandingkan dengan beberapa metode klasifikasi *machine learning* lainnya karena dapat mengenali pola dan menganalisis data dengan dimensi tinggi sehingga tingkat akurasi yang dihasilkan lebih baik (Ali Nur Ikhsan et al., 2025).

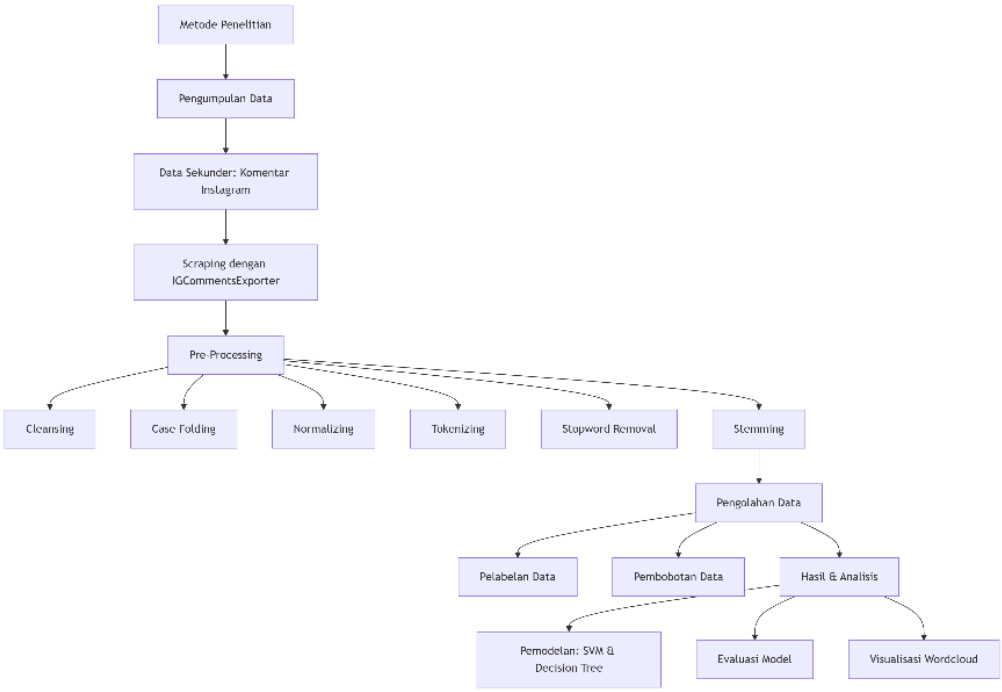
Penelitian serupa yang sudah dilakukan menyatakan bahwa analisis sentimen terhadap program makan bergizi menggunakan dua algoritma yakni *Support Vector Machine* dan *Random Forest*, menunjukkan sentimen negatif yang mendominasi dengan akurasi tertinggi sebesar 85,74% dengaan menggunakan algoritma SVM (Triningsih, 2025). Penelitian lainnya juga menunjukkan bahwa Algoritma SVM terbukti mampu mengelola data dengan baik, menghasilkan nilai akurasi yang tinggi, yaitu sebesar 94% (Purwanti & Sugiyono, 2024). Penelitian lebih jauh juga menunjukkan bahwa hasil perbandingan model klasifikasi *Support Vector Machine* (SVM) lebih unggul dibandingkan *Naïve Bayes Classifier* (NBC) dalam hal akurasi, presisi, recall, dan F1-score, meskipun SVM membutuhkan waktu komputasi yang lebih lama (Ramadhani, 2024). Penelitian ini penting karena kajian analisis sentimen terhadap program makan bergizi gratis di media sosial Instagram masih terbatas, khususnya dengan pendekatan algoritma SVM dan *Decision Tree*. Dengan memanfaatkan data komentar dari unggahan bertagar #MakanBergiziGratis, penelitian ini bertujuan memberikan gambaran mengenai persepsi masyarakat di Instagram terhadap realisasi program makan bergizi gratis.

Hasil analisis diharapkan dapat menjadi masukan berharga bagi para pembuat kebijakan dalam mengevaluasi dan memperbaiki implementasi program untuk depan.

2 METODE

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis algoritma *Support Vector Machine* dan *Decision Tree*. Pendekatan ini dipilih karena sentimen karena kinerja dan efisiensinya dalam memproses data dalam jumlah besar, termasuk teks media sosial (Yuspita & Suryono, 2023). Desain penelitian ini dirancang untuk menganalisis tanggapan masyarakat terhadap implementasi Program Makan Siang Gratis yang disampaikan melalui komentar pada unggahan di media sosial Instagram. Penelitian ini terdiri dari beberapa tahapan, dimulai dari pengumpulan data komentar melalui proses *scraping*, dilanjutkan dengan tahap praproses data, ekstraksi fitur, pelatihan model menggunakan algoritma SVM dan *Decision Tree*, serta evaluasi kinerja model berdasarkan metrik akurasi, presisi, *recall*, dan f1-*score*. Seluruh proses analisis dilakukan dengan memanfaatkan bahasa pemrograman Python yang dijalankan melalui platform Google Colab, karena platform ini mendukung komputasi berbasis *cloud* yang efisien.



Gambar 1. Alur Metode Penelitian

2.2 Tahap Pengumpulan Data

Penelitian ini menggunakan data sekunder berupa komentar masyarakat pada unggahan Instagram yang membahas Program Makan Siang Gratis. Data dikumpulkan melalui metode scraping menggunakan ekstensi Google Chrome bernama *IGCommentsExpor* (Arrafiq & Kurniawan, 2024), sebuah ekstensi Google Chrome yang memungkinkan pengguna untuk mengekstraksi komentar Instagram dalam jumlah besar. Setelah data dikumpulkan, file tersebut disimpan dalam bentuk .csv agar dapat diolah secara otomatis. Adapun teknik pengambilan sampel menggunakan metode *purposive sampling*, yakni dengan memilih unggahan yang secara spesifik membahas program dimaksud. Pengumpulan data dilakukan pada periode 6 Januari 2024 hingga 27 April 2025, dengan unggahan yang dipilih memiliki tagar #makanbergizigratis. Setelah dikumpulkan, data tersebut melalui tahapan praproses untuk menjadikannya lebih terstruktur dan siap digunakan dalam analisis sentimen.

2.3 Tahap Pre-Processing

Tahapan praproses data bertujuan untuk memperoleh data yang bersih dan terstruktur, sehingga memudahkan proses pengolahan pada tahap selanjutnya (Saputra & Hasan, 2024). Tahapan dalam *text preprocessing* meliputi *cleansing, case folding, normalizing, tokenizing, stopword removal,* dan *stemming*. Penjelasan dari masing-masing tahapan adalah sebagai berikut:

a. *Cleansing*: Menghapus karakter tidak diperlukan seperti tanda baca, emoji, angka, URL, *username, mention*, spasi berlebih, dan *hashtag*.
b. *Case Folding*: Mengubah seluruh huruf menjadi kecil untuk menyeragamkan teks (Anggriyani & Fakhriza, 2024).
c. *Normalizing*: Mengubah kata tidak baku atau singkatan menjadi bentuk baku sesuai KBBI.
d. *Tokenizing*: Memisahkan kalimat menjadi kata-kata (token) berdasarkan spasi atau tanda baca (Rahmani & Sejati, 2024).
e. *Stopword Removal*: Menghapus kata-kata yang tidak relevan menggunakan daftar stopword dari pustaka Sastrawi.
f. *Stemming*: Menghilangkan imbuhan untuk mengembalikan kata ke bentuk dasarnya (Sitanggang et al., 2024).

2.4 Tahap Pengolahan Data

Tahapan ini merupakan proses mengelola, memahami, dan mengkategorikan data teks. Adapun tahapannya sebagai berikut.

a. Pelabelan Data

Setelah melalui tahap *pre-processing*, langkah berikutnya yaitu memeberikan label sentimen positif dan negatif pada setiap komentar. Proses pelabelan ini akan membantu dalam pemodelan dan analisis berikutnya.

b. Pembobotan Data

Setelah melakukan pelabelan data, proses selanjutnya yang dilakukan adalah pembobotan data dengan cara *Term Frequency-Inverse Document Frequency* (TD-IDF) pada data teks. *Term Frequency* (TF) menghitung bobot kata berdasarkan seberapa sering kata tersebut muncul, sedangkan *Inverse Document Frequency* (IDF) mengukur relevansi kata dalam dokumen dengan memprioritaskan kata-kata yang jarang muncul. Nilai IDF semakin tinggi jika kata lebih unik. Metode ini banyak digunakan karena dapat menghasilkan analisis yang lebih akurat dan efisien, terutama untuk keperluan klasifikasi, analisis sentimen, dan pencarian web. Pemilihan TF-IDF sebagai metode ekstraksi kata membantu mengurangi pengaruh kata-kata yang sering muncul dan lebih menekankan kata-kata yang lebih penting serta relevan untuk analisis sentimen. Rumus untuk menghitung *Term Frequency* (TF) adalah (Vebrian & Kustiyono, 2025):

$$TF = \frac{\text{Jumlah kemunculan term dalam data}}{Total\ jumlah\ termn\ dalam\ data}$$

Apabila nilai Term Frequency(TF) dari data tersebut sudah didapatkan berikutnya ialah menentukan nilai Inverse Document Frequency (IDF) terlebih dahulu menggunakan persamaan (Septiani & Isabela, 2023):

$$IDF = \log(\frac{N}{df})$$

N dipersamaan tersebut melambangkan jumlah keseluruhan data tersebut. Langkah berikutnya ialah mengalikan hasil dari persamaan TF dan IDF, sehingga ditemukan hasil TF-IDF seutuhnya. Nilai atau hasil TF-IDF didapatkan dengan menggunakan persamaan diberikut ini:

$$\text{TD} - \text{IDF} = \text{TF} \times \text{IDF}$$

2.5 Tahap Hasil dan Analisis

a. Pemodelan Data

Support Vector Machine (SVM) adalah algoritma klasifikasi yang mencari hyperplane optimal untuk memisahkan dua kelas data dengan margin maksimum. Ada beberapa teorema dan rumus kunci yang mendasari cara kerja SVM. Adapun teorema rumusnya adalah:

Fungsi keputusan SVM:

$$h(x) \ = \ sign(w^T \ x \ + \ b)$$

Di mana x adalah vektor fitur (misal TF-IDF), w vektor bobot, dan b bias.

Optimasi margin maksimum:

$$\min_{w,b} \frac{1}{2} \ \|w\|^2 + \ \mathrm{C} \ \sum_{j=1}^{N} max(0,1 \ - \mathrm{yj}(w^T xj \ + b) \ )$$

dengan $yj \ \epsilon \ \{+1, -1\}$ sebagai label sentimen (positif/negatif).

Kernel trick (Ha et al., 2011):

$$h(x) = sign\left(\sum_{n=1}^{N} \beta_n K(x_n, x) + b\right)$$

Sedangkan *Decision Tree* merupakan algoritma klasifikasi yang membangun model dalam bentuk struktur pohon (Purnamawati et al., 2020), di mana setiap node internal merepresentasikan pengujian terhadap atribut tertentu, setiap cabang menunjukkan hasil dari pengujian tersebut, dan setiap *leaf node* menggambarkan keputusan atau hasil klasifikasi. Dalam proses pembentukan pohon keputusan, dua konsep matematika utama yang digunakan adalah *entropy* dan *information gain; entropy* mengukur tingkat ketidakpastian atau *impurity* dalam dataset, sedangkan information gain menghitung pengurangan entropy setelah dataset dibagi berdasarkan atribut tertentu. Atribut dengan nilai information gain tertinggi dipilih sebagai node percabangan untuk membagi data secara bertahap hingga diperoleh klasifikasi yang optimal. Adapun teorema dari algoritma Decision Tree adalah:

Entropy untuk mengukur ketidakpastian kelas sentimen:

$$Entropy(S) = \ - \sum_{i=1}^{m} P_i \log_2(P_i)$$

Information Gain untuk memilih atribut (fitur kata) terbaik (Ha et al., 2011):

$$Gain(S, A) = Entropy(S) \ - \sum_{v \epsilon Values(A)} \frac{|S_v|}{|S|} \ Entropy \ (S_v)$$

b. Evaluasi

Hasil klasifikasi dari kedua model akan dianalisis untuk menilai performa masing-masing dalam mengklasifikasikan sentimen positif dan negatif. Evaluasi dilakukan dengan mengukur metrik kinerja seperti akurasi, presisi, *recall*, dan F1-score dari hasil prediksi kedua model (Riswan et al., 2025). Perbandingan hasil ini bertujuan untuk menentukan model mana yang lebih efektif dan akurat dalam mengklasifikasikan sentimen terkait data yang digunakan. Analisis hasil evaluasi juga akan memberikan gambaran sejauh mana masing-masing model mampu mengenali pola sentimen positif dan negatif, serta bagaimana karakteristik data mempengaruhi performa kedua algoritma tersebut.

c. Visualisasi Wordcloud

Hasil dari proses klasifikasi sentimen divisualisasikan menggunakan pustaka *Wordcloud. Word cloud* merupakan bentuk visualisasi data yang menampilkan kata-kata yang paling sering muncul dalam kumpulan data, di mana ukuran masing-masing kata mencerminkan frekuensi kemunculannya (Nursinggah et al., 2024).

3 HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Hasil awal *scrapping* data dengan menggunakan google chrome bernama *IGCommentsExport,* diperoleh 2433 komentar dari 52 unggahan dengan hastag #MakanBergiziGratis sejak 6 januari 2024 hingga 27 April 2025. Namun, setelah melakukan proses penyaringan untuk menghapus data yang tidak berguna atau data yang tidak bersifat sentimen dengan memanfaatkan memanfaatkan bahasa pemrograman python, maka data yang akan diolah hanya 2327 komentar. Data tersebut disimpan dalam bentuk csv.

Tabel 1. Hasil Penyaringan Data

Kategori	Jumlah
Sentimen	2327
Komentar promosi/spam	47
Komentar berisi emoji saja	47
Komentar kosong	11
Komentar tidak mengandung opini	1
Jumlah komentar sentimen yang siap diproses	2433

3.2 *Pre-processing* Data

Setelah data disimpan dalam bentuk csv, selanjutnya akan dilakukan tahap preprocessing data yaitu dilakukan dengan mengolah data tidak terstruktur menjadi data yang lebih terstruktur agar tidak menyebabkan berkurangnya tingkat akurasi pada hasil analisis. Adapun tahapan data *preprocessing* yaitu *cleaning, case folding, normalisasi, tokenizing, filtering,* dan *stemming.*

Tabel 2. Contoh Hasil *Pre-processing* data

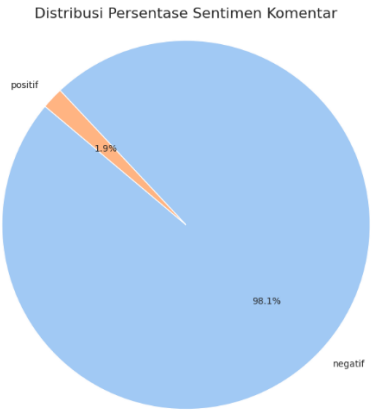
3.3 Proses Pelabelan

Pemberian label ini penting untuk menyediakan target (*ground truth)* yang akan digunakan oleh model klasifikasi. Dalam penelitian ini, proses *labelling* sentimen dilakukan secara otomatis menggunakan pustaka *TextBlob.* Pustaka *TextBlob* menyediakan fungsi analisis sentimen yang menghasilkan skor polaritas. Skor polaritas berkisar antara -1.0 (sangat negatif) hingga +1.0 (sangat positif), dengan 0.0 menunjukkan netral. Dari hasil pelabelan data tersebut, diperoleh komentar pada sosial media Instagram sebanyak 44 sentimen Positif (1,9 %) dan 2283 sentimen Negatif (98,9 %).

Tabel 3. Contoh Hasil Proses Pelabelan Data

Teks Komentar	Hasil Sentimen
program tidak jelas	Negatif
tuju benar program bagus kalau tidak tergesagesa terlalu paksa benar mungkin mulai lingkup lebih kecil contoh daerah memang angka stunting gizi buruk tingi fokus anak taman kanak-kanak sekolah dasar tahap kalau tidak perlu efisiensi semua jalan tanpa mengangu	Negatif
ladang korupsi	Negatif

Proses	Hasil
Data awal	19302664785,roy.suyatno,18041016554375363,”Setuju sekali, sebenarnya program ini bagus kalo tidak tergesa-gesa dan terlalu memaksakan sebenarnya. Mungkin bisa dimulai dari lingkup yang lebih kecil dulu . Contoh: untuk daerah2 yang memang angka stunting atau gizi buruk tinggi, fokus ke anak TK atau SD saja duluðŸ™Œ Bertahaplah kalo bisa jadi tidak perlu efisiensi sana sini, semua bisa jalan tanpa menggangu yang lainðŸ™Œ”,https://www.instagram.com/roy.suyatno,https://instagram.fsrg6-1.fna.fbcdn.net/v/t51.2885-19/485836430_1377411873278474_6256670242731989090_n.jpg?stp=dst-jpg_s150x150_tt6&_nc_ht=instagram.fsrg6-1.fna.fbcdn.net&_nc_cat=102&_nc_oc=Q6cZ2QFHvX2S2A2Ihs5nt9osP86Fx0x4MDeVsmDXMJRsxDal69C6dzg1jL612jMZg-_5LvU&_nc_ohc=jM0HzYZwn0wQ7kNvwFSm6cr&_nc_gid=sYdTiU_92MvrucDT8Q90ug&edm=AI-cjbYBAAAA&ccb=7-5&oh=00_AfE4Qnlw_UBSHvjCduZy1bBIrqT8rq4b8ELQCNWovHH7jA&oe=68137424&_nc_sid=0985b0,"26/2/2025, 08.55.52",sentimen
Cleaning	Setuju sekali sebenarnya program ini bagus kalo tidak tergesagesa dan terlalu memaksakan sebenarnya Mungkin bisa dimulai dari lingkup yang lebih kecil dulu Contoh untuk daerah yang memang angka stunting atau gizi buruk tinggi fokus ke anak TK atau SD saja dulu Bertahaplah kalo bisa jadi tidak perlu efisiensi sana sini semua bisa jalan tanpa menggangu yang lain
Case folding	setuju sekali sebenarnya program ini bagus kalo tidak tergesagesa dan terlalu memaksakan sebenarnya mungkin bisa dimulai dari lingkup yang lebih kecil dulu contoh untuk daerah yang memang angka stunting atau gizi buruk tinggi fokus ke anak tk atau sd saja dulu bertahaplah kalo bisa jadi tidak perlu efisiensi sana sini semua bisa jalan tanpa menggangu yang lain
Normalizing	setuju sekali sebenarnya program ini bagus kalau tidak tergesagesa dan terlalu memaksakan sebenarnya mungkin bisa dimulai dari lingkup yang lebih kecil dulu contoh untuk daerah yang memang angka stunting atau gizi buruk tingi fokus ke anak taman kanak-kanak atau sekolah dasar saja dulu bertahaplah kalau bisa jadi tidak perlu efisiensi sana sini semua bisa jalan tanpa mengangu yang lain
Tokenizing	[setuju, sekali, sebenarnya, program, ini, bagus, kalau, tidak, tergesagesa, dan, terlalu, memaksakan, sebenarnya, mungkin, bisa, dimulai, dari, lingkup, yang, lebih, kecil, dulu, contoh, untuk, daerah, yang, memang, angka, stunting, atau, gizi, buruk, tingi, fokus, ke, anak, taman, kanak-kanak, atau, sekolah, dasar, saja, dulu, bertahaplah, kalau, bisa, jadi, tidak, perlu, efisiensi, sana, sini, semua, bisa, jalan, tanpa, mengangu, yang, lain]
Stopword	[setuju, sebenarnya, program, bagus, kalau, tidak, tergesagesa, terlalu, memaksakan, sebenarnya, mungkin, dimulai, lingkup, lebih, kecil, contoh, daerah, memang, angka, stunting, gizi, buruk, tingi, fokus, anak, taman, kanak-kanak, sekolah, dasar, bertahaplah, kalau, tidak, perlu, efisiensi, semua, jalan, tanpa, mengangu]
Stemming	tuju benar program bagus kalau tidak tergesagesa terlalu paksa benar mungkin mulai lingkup lebih kecil contoh daerah memang angka stunting gizi buruk tingi fokus anak taman kanak-kanak sekolah dasar tahap kalau tidak perlu efisiensi semua jalan tanpa mengangu



Gambar 2. Visualisasi *Pie Chart* Hasil Pelabelan Data

3.4 Proses Pembobotan Data TD IDF

Berdasarkan hasil ekstraksi, diperoleh matriks TF-IDF berukuran 2327×4814, yang mencerminkan 2327 dokumen dan 4814 fitur (kata unik hasil dari tahap pra-pemrosesan). Transformasi ini menghasilkan vektor numerik di mana setiap nilai menunjukkan tingkat kepentingan suatu kata dalam konteks sebuah komentar dibandingkan dengan keseluruhan korpus. Berikut contoh dari tahapan pembobotan pada beberapa data.

Tabel 4. Contoh Hasil Pembobotan Data

No.	Anak	Gizi	Gratis	Makan
1	0	0	0	0
2	0	0	0	0
3	1	0	0	0
4	0	0,654	0	0,757
5	0,671	0	0,741	0
6	0	0	0	0
7	0,849	0,523	0	0

3.5 Klasifikasi Algoritma *Support Vector Machine* dan *Decision Tree*

3.5.1 Pembagian Data

Tahapan selanjutnya adalah membagi dataset sebagai pelatihan dan tahapan pengujian model. Pada penelitian ini, data dibagi dengan rasio 80:20, yang artinya 80% data diolah untuk pelatihan model dan 20% data sisa digunakan untuk pengujian kinerja model. Rasio pembagian ini merupakan pendekatan yang lazim digunakan untuk memastikan bahwa model memperoleh data yang memadai untuk proses pelatihan sekaligus memiliki data yang cukup untuk pengujian. Berdasarkan rasio tersebut, diperoleh data training sebanyak 1.861 dan data testing sebanyak 466 (Azhari & Haryanto, 2024).

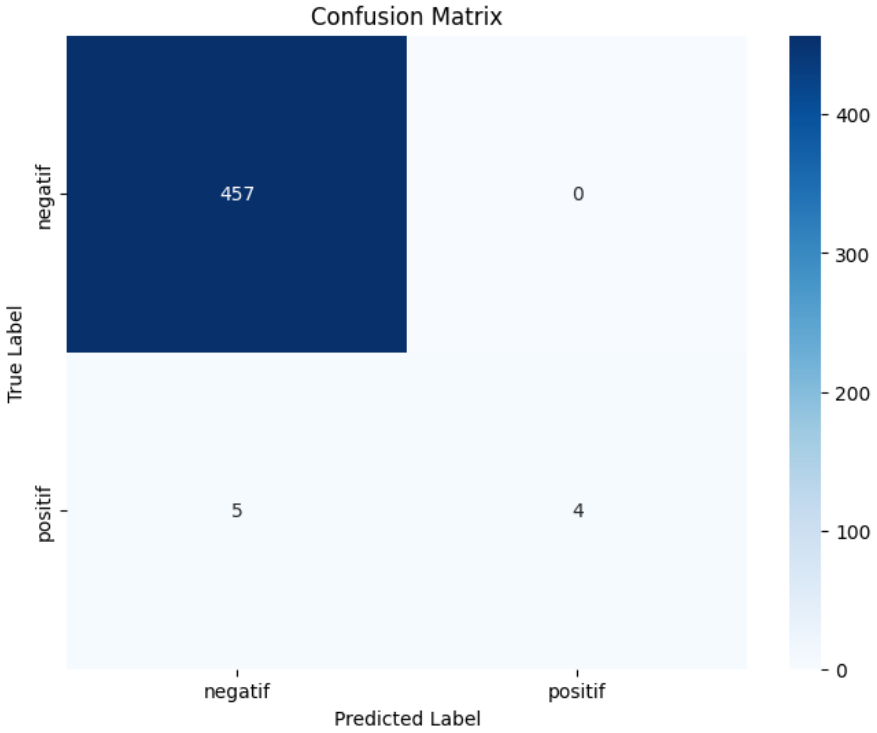
a. *Support Vector Machine*

Model *Support Vector Machine* (SVM) pada penelitian ini diinisialisasi dengan kernel linear (kernel='linear') dan pengaturan bobot kelas secara seimbang (*class_weight='balanced'*). Evaluasi kinerja model dilakukan dengan menggunakan metrik klasifikasi standar, yaitu *Accuracy*, *Precision*, *Recall*, dan F1-*score*, serta *Confusion Matrix*.

Tabel 5. Klasifikasi *Support Vector Machine*

	Precision	Recall	F1-score	Support
Negatif	0.99	1.00	0.99	457

Positif	1.00	0.44	0.62	9
Accuracy			0.99	466
Macro avg	0.99	0.72	0.80	466
Weiighted avg	0.99	0.99	0.99	466



Gambar 3. *Confusion Matrix Support Vector Machine*

Berdasarkan hasil klasifikasi dan *confusion matrix*, model menunjukkan tingkat akurasi sebesar 0,99 atau 99%, yang mengindikasikan bahwa model memiliki kemampuan klasifikasi yang tinggi secara umum terhadap data uji. Hasil menunjukkan bahwa model memiliki kinerja yang sangat tinggi pada kelas negatif dengan *precision* dan *recall* masing-masing sebesar 0,99 dan 1,00, menghasilkan nilai f1-*score* sebesar 0,99 dari total 457 sampel. Namun, pada kelas positif, kinerja model masih terbatas dengan recall hanya sebesar 0,44 meskipun *precision* mencapai 1,00. Hal ini menunjukkan bahwa dari seluruh komentar yang sebenarnya positif, hanya sebagian kecil yang berhasil dikenali dengan benar oleh model, sementara semua komentar yang diprediksi sebagai positif benar-benar merupakan komentar positif (tidak ada *False Positive*). *Confusion matrix* memperkuat temuan ini, di mana sebanyak 5 dari 9 komentar positif salah diklasifikasikan sebagai negatif. Ketidakseimbangan jumlah data antar kelas kemungkinan besar berkontribusi terhadap rendahnya *recall* pada kelas minoritas.

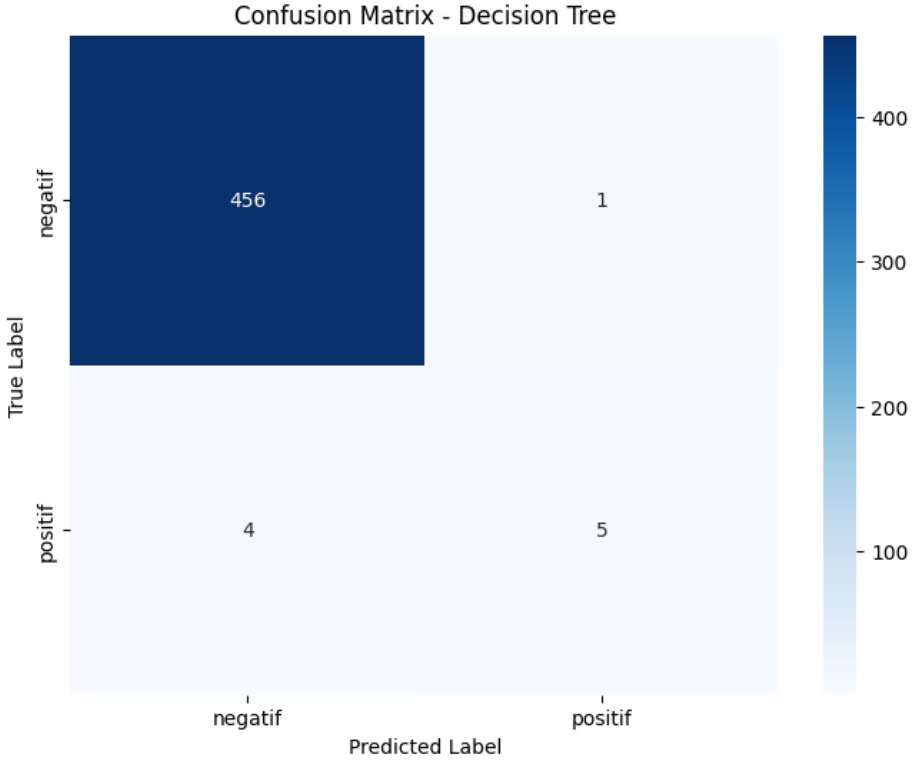
b. *Decision Tree*

Model *Decision Tree* pada penelitian ini dikonstruksi menggunakan pustaka *Scikit-learn* dengan kelas *DecisionTreeClassifier* dan nilai random_state=42 guna menjamin reprodusibilitas hasil. Pada tahap awal, model dijalankan menggunakan parameter default, tanpa melakukan penyetelan terhadap parameter-parameter penting seperti *criterion, max_depth*, dan *min_samples_split*. Evaluasi model dilakukan melalui metrik-metrik klasifikasi standar: *accuracy, precision, recall*, dan f1-*score*, serta divisualisasikan melalui *confusion matrix.*

Tabel 7. Klasifikasi *Decision Tree*

	Precision	Recall	F1-score	Support
Negatif	0.99	1.00	0.99	457
Positif	0.83	0.56	0.67	9
Accuracy			0.99	466

Macro avg	0.99	0.78	0.83	466
Weighted avg	0.99	0.99	0.99	466



Gambar 4. *Confusion Matrix Decision Tree*

Berdasarkan *classification report* dan *confusion matrix* yang ditampilkan, model *Decision Tree* menghasilkan tingkat akurasi sebesar 0,99 atau 99%, menunjukkan performa klasifikasi yang tinggi secara keseluruhan. Untuk kelas negatif, model mencatat *precision* dan *recall* yang sangat baik, masing-masing sebesar 0,99 dan 1,00, serta f1-*score* sebesar 0,99 dari 457 sampel, dengan hanya satu kesalahan klasifikasi (*False Positive*). Sementara itu, pada kelas positif, meskipun *precision* relatif baik sebesar 0,83, recall hanya mencapai 0,56, menghasilkan f1-*score* sebesar 0,67 dari 9 sampel positif. Hal ini menunjukkan bahwa meskipun model cukup berhati-hati dalam menetapkan label positif (*precision* tinggi), kemampuannya dalam mendeteksi seluruh komentar positif masih terbatas.

Confusion matrix memperlihatkan bahwa dari 9 komentar positif, 5 diklasifikasikan dengan benar, sementara 4 lainnya salah dikenali sebagai negatif. Metrik *macro average* dan *weighted average* berturut-turut sebesar 0,83 dan 0,99 mencerminkan bahwa performa model terhadap kelas minoritas positif belum setara dengan kelas mayoritas negatif. Hasil ini menegaskan bahwa meskipun akurasi tinggi dapat tercapai, evaluasi berbasis metrik per kelas tetap penting dalam konteks klasifikasi sentimen, terutama ketika distribusi kelas tidak seimbang.

3.5.2 Evaluasi Perbandingan Kinerja Model SVM dan *Decision Tree*

Dari segi akurasi, kedua model menunjukkan performa yang sangat tinggi, dengan nilai akurasi identik sebesar 0,99. Namun, ketika ditinjau lebih dalam melalui metrik-metrik klasifikasi per kelas, terdapat perbedaan performa yang cukup mencolok terutama pada penanganan kelas minoritas, yakni sentimen positif. Model SVM menghasilkan nilai *precision* sebesar 1,00 namun dengan *recall* yang rendah yaitu 0,44 dan f1-*score* sebesar 0,62 pada kelas positif. Hal ini mengindikasikan bahwa meskipun semua prediksi positif benar (tidak terjadi *False Positive*), model SVM gagal mengidentifikasi sebagian besar komentar yang sebenarnya positif (terjadi 5 *False Negative* dari total 9). Sebaliknya, model *Decision Tree* menghasilkan precision yang sedikit lebih rendah untuk kelas positif 0,83, namun dengan peningkatan recall

menjadi 0,56 dan f1-*score* sebesar 0,67. Meskipun masih terdapat kesalahan klasifikasi, model *Decision Tree* lebih seimbang dalam mendeteksi komentar positif dibandingkan SVM.

Kedua model memiliki kinerja yang hampir sempurna dalam mengklasifikasikan komentar negatif, dengan f1-*score* sebesar 0,99 dan *recall* sebesar 1,00. Ini menunjukkan dominasi kelas negatif yang cukup kuat dalam dataset dan keberhasilan kedua model dalam mengenali kelas mayoritas. Akan tetapi, karena distribusi data yang tidak seimbang, akurasi tinggi tidak cukup merepresentasikan performa keseluruhan model secara adil. Oleh karena itu, analisis berdasarkan *precision, recall,* dan f1-*score* per kelas menjadi penting.

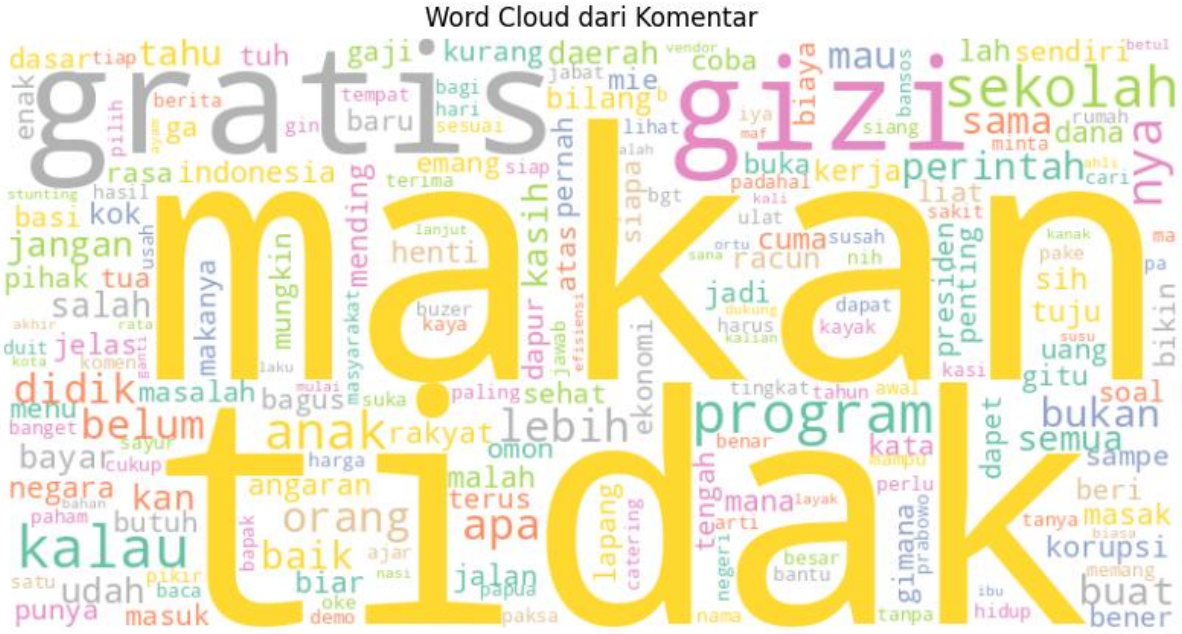
Secara keseluruhan, SVM menunjukkan keunggulan dalam menghindari prediksi positif palsu *(false positive)*, namun cenderung abai terhadap data positif yang sebenarnya (*false negative* tinggi), sedangkan Decision Tree menawarkan keseimbangan yang lebih baik antara *precision* dan *recall* pada kelas minoritas, meskipun tetap mengalami keterbatasan. Dengan demikian, pemilihan model ideal bergantung pada prioritas tugas klasifikasi: jika tujuan utamanya adalah meminimalkan kesalahan positif palsu, maka SVM lebih sesuai; tetapi jika diperlukan kemampuan deteksi lebih baik terhadap kelas minoritas, maka pendekatan *Decision Tree* dapat dipertimbangkan sebagai alternatif yang lebih adaptif.

Tabel 7. Perbandingan Hasil Klasifikasi

	SVM	Decision Tree
Accuracy	0.99	0.99
Precision (Negatif)	0.99	0.99
Recall (Negatif)	1.00	1.00
F1-Score (Negatif)	0.99	0.99
Precision (Positif)	1.00	0.83
Recall (Positif)	0.44	0.56
F1-Score (Positif)	0.62	0.67

3.5.3 Visualisasi Data Sentimen

Visualisasi data dilakukan untuk menangkap topik-topik utama yang kerap dibicarakan oleh pengguna dalam komentar mereka. Dengan menggunakan teknik *Word Cloud*, kata-kata yang paling sering muncul divisualisasikan berdasarkan frekuensi kemunculannya dalam korpus data.



Gambar 4. Hasil Visualisasi Sentimen

Pada Gambar di atas, kata-kata yang ditampilkan dengan ukuran huruf lebih besar menandakan frekuensi kemunculannya yang lebih tinggi dalam kumpulan data. Beberapa kata dominan seperti *makan, tidak, gratis, program, gizi,* dan *sekolah* menunjukkan bahwa topik diskusi utama berpusat pada isu penyediaan makanan, manfaat gizi, dan implementasi program di lingkungan pendidikan.

Kemunculan kata *tidak* dalam ukuran yang sangat besar mengindikasikan banyaknya komentar yang mengandung penyangkalan, kekhawatiran, atau kritik terhadap program tersebut. Di sisi lain, kata *gratis* dan *gizi* memperlihatkan bahwa aspek ketersediaan makanan tanpa biaya serta kualitas gizi menjadi perhatian utama masyarakat. Kata *program*, *sekolah*, dan *didik* juga banyak disebut, memperkuat dugaan bahwa audiens mengaitkan kebijakan ini erat dengan sektor pendidikan, khususnya sekolah dasar hingga menengah.

Selain itu, munculnya kata-kata seperti *korupsi*, *rakyat*, *anggaran*, dan *pemerintah* mencerminkan adanya kekhawatiran publik terhadap tata kelola dana dan transparansi pelaksanaan program. Kata-kata seperti *baik*, *bantu*, dan *cukup* menunjukkan sebagian sentimen positif atau apresiasi. Dengan demikian, *Word Cloud* tidak hanya berfungsi sebagai alat eksplorasi data awal, tetapi juga memperkuat temuan klasifikasi sentimen dengan menyajikan konteks kata yang beragam dari opini masyarakat.

4 KESIMPULAN

Penelitian ini berhasil menganalisis sentimen masyarakat terhadap Program Makan Bergizi Gratis melalui komentar di Instagram menggunakan algoritma *Support Vector Machine* dan *Decision Tree*. Kedua model menunjukkan akurasi tinggi sebesar 99%. Namun, terdapat perbedaan performa dalam menangani kelas minoritas, dengan SVM lebih unggul dalam presisi (1,00) tetapi lemah dalam recall (0,44) terhadap sentimen positif, sedangkan *Decision Tree* menunjukkan performa lebih seimbang dengan presisi (0,83) dan recall (0,56). Temuan ini sekaligus mengungkap keterbatasan penelitian, terutama terkait ketidakseimbangan data yang sangat timpang (98,1% negatif dengan 1,9% positif) yang menyebabkan model cenderung bias terhadap kelas mayoritas. Metode pelabelan otomatis menggunakan *TextBlob* juga berpotensi kurang akurat dalam menangkap nuansa bahasa seperti ironi atau konteks ambigu. Visualisasi *word cloud* mengonfirmasi bahwa diskusi publik tidak hanya berfokus pada aspek gizi, tetapi juga menyoroti isu implementasi, transparansi anggaran, dan kekhawatiran akan korupsi. Hasil penelitian ini diharapkan dapat berkontribusi bagi pengambil kebijakan dalam memahami persepsi publik secara lebih komprehensif. Untuk penelitian selanjutnya, disarankan penggunaan dataset yang lebih besar dan seimbang, eksplorasi media sosial lain, serta penerapan algoritma berbasis *deep learning* seperti LSTM atau BERT untuk meningkatkan performa klasifikasi, khususnya pada kelas minoritas.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah mendukung penelitian ini. Terutama kepada Universitas Islam Negeri K.H Abdurrahman Wahid Pekalongan yang telah memberikan fasilitas dan dukungan akademik. Ucapan terima kasih juga disampaikan kepada rekan-rekan peneliti dan kolega yang telah memberikan masukan berharga selama proses penelitian. Tidak lupa, kami berterima kasih kepada para responden yang telah berkontribusi melalui komentar di Instagram, sehingga data penelitian ini dapat terkumpul dengan baik. Dukungan moril dan spiritual dari keluarga serta teman-teman juga sangat berarti dalam menyelesaikan penelitian ini. Semoga kontribusi semua pihak mendapat balasan yang setimpal.

DAFTAR PUSTAKA

Ali Nur Ikhsan, Pungkas Subarkah, Alifah Dafa Iftinani, & Fadilah, A. N. (2025). Analisis

Sentimen Kenaikan PPN menggunakan Algoritma Naïve Bayes dan Support Vector Machine. *INFOTEKMESIN Media Informasi Untuk Pengembangan Peneliti Teknik, 16*(1), 181–189. https://doi.org/10.35970/infotekmesin.v16i1.2518

Anggriyani, W., & Fakhriza, M. (2024). Analisis Sentimen Program Makan Gratis Pada Media Sosial X Menggunakan Metode NLP. *Journal of Computer System and Informatics (JoSYC), 5*(4), 1033–1042. https://doi.org/10.47065/josyc.v5i4.5826

Arrafiq, M. S., & Kurniawan, R. (2024). Analisis Sentimen Pengguna Terhadap Layanan Aplikasi Seabank Indonesia di Instagram Menggunakan Metode Support Vector Machine. *Journal of Information System …*, *5*(4), 1280–1291. https://doi.org/10.47065/josh.v5i4.5620

Azhari, I. C., & Haryanto, T. (2024). Modeling Of Hyperparameter Tuned RNN-LSTM and Deep Learning For Garlic Price Forecasting In Indonesia. *Journal of Informatics and Telecommunication Engineering*, *7*(2), 502–513. https://doi.org/10.31289/jite.v7i2.10714

Ha, J., Kambe, M., & Pe, J. (2011). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. https://doi.org/10.1016/C2009-0-61819-5

Lestari, D. L., Gusmira, Y. H., Mashdarul, A., & Amelia, A. Y. (2024). Free Nutritious Meal Policy As A Solution To Overcoming The Stunting Problem In Indonesia. *INNOVATIVE: Journal Of Social Science Research Volume*, *4*(4), 10021–10031.

Merlinda, A. A., & Yusuf, Y. (2025). Analisis Program Makan Gratis Prabowo Subianto Terhadap Strategi Peningkatan Motivasi Belajar Siswa di Sekolah Tinjauan dari Perspektif Sosiologi Pendidikan. *Ranah Research : Journal of Multidisciplinary Research and Development*, *7*(2), 1364–1373. https://doi.org/https://doi.org/10.38035/rrj.v7i2

Nursinggah, L., Ruuhwan, R., & Mufizar, T. (2024). Analisis Sentimen Pengguna Aplikasi X Terhadap Program Makan Siang Gratis Dengan Metode Naïve Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3). https://doi.org/10.23960/jitet.v12i3.4336

Pambudi, A. W. (2024). Indonesia Government ' s Efforts to Eradicate Stunting Through the Free Nutritious Meal Program for Children : Opportunities and Challenges. *Jurnal Politikom Indonesiana: Kajian Ilmu Pemerintahan, Ilmu Politik Dan Ilmu Komunikasi*, *9*(2), 157–182.

Purnamawati, A., Nugroho, W., Putri, D., & Hidayat, W. F. (2020). Deteksi Penyakit Daun pada Tanaman Padi Menggunakan Algoritma Decision Tree, Random Forest, Naïve Bayes, SVMdan KNN. *InfoTekJar: Jurnal Nasional Informatika Dan Teknologi Jaringan*, *5*(1), 212–215.

Purwanti, Z., & Sugiyono. (2024). Pemodelan Text Mining untuk Analisis Sentimen Terhadap Program Makan Siang Gratis di Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, *5*(3), 3065–3079. https://doi.org/https://doi.org/10.35870/jimik.v5i3.1001

Putriyekti, A., Sulistiawati, D., Khairani, N., & Atmaja, R. R. (2025). ANALISIS MULTIDIMENSIONAL SENTIMEN MASYARAKAT TERHADAP PROGRAM MAKAN BERGIZI GRATIS PADA MEDIA SOSIAL X. *Integrative Perspectives of Social and Science Journal (IPSSJ)*, *2*(1), 1004–1024.

Rahmani, B., & Sejati, N. (2024). Model Prediksi Tipe Komentar Di Instagram Terkait Pemilu 2024 Menggunakan Metode Winnowing. *Jutisi: Jurnal Ilmiah Teknik Informatika*, 2414–2425.

Ramadhani, A. (2024). ANALISIS SENTIMEN TANGGAPAN PUBLIK TERHADAP PROGRAM MAKAN SIANG GRATIS MENGGUNAKAN ALGORITMA NBC DAN SVM TUGAS [Universitas Islam Negeri Sultan Syarif Kasim Riau]. In *Building of Informatics, Technology and Science (BITS)* (Vol. 6, Issue 3). https://doi.org/10.47065/bits.v6i3.6188

Riswan, M., Primajaya, A., Susilo, A., & Irawan, Y. (2025). ANALISIS SENTIMEN

TERHADAP PEMBERITAAN HASIL REKAPITULASI PEMILU PRESIDEN 2024 PADA MEDIA SOSIAL INSTAGRAM MENGGUNAKAN NAIVE BAYES CLASSIFIER. *JITET (Jurnal Informatika Dan Teknik Elektro Terapan)*, *13*(1), 203–211. https://doi.org/http://dx.doi.org/10.23960/jitet.v13i1.5559

Ritonga, A. R., & Sazali, H. (2025). ANALISIS KOMUNIKASI PEMBANGUNAN TERHADAP REGULASI DAN KEBIJAKAN PROGRAM MAKAN BERGIZI GRATIS. *Vijnana: Jurnal Hasil Penelitian Multidisiplin*, *1*(1), 32–40.

Saleh, M. F., & Imanda, R. (2025). Public Sentiment Analysis of the Free Meal Program : A Comparison of Naive Bayes and Support Vector Machine Methods on the Twitter (X) Social Media Platform. *Journal of Applied Informatics and Computing (JAIC)*, *9*(1), 131–139.

Saputra, R., & Hasan, F. N. (2024). Analisis Sentimen Terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naive Bayes. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *6*(3), 411–419. https://doi.org/10.47233/jteksis.v6i3.1378

Septiani, D., & Isabela, I. (2023). Analisis Term Frequency Inverse Document Frequency (TF-IDF) Dalam Temu Kembali Informasi Pada Dokumen Teks. *SINTESIA: Jurnal Sistem Dan Teknologi Informasi Indonesia*, *1*(2), 81–88.

Sitanggang, A., Umaidah, Y., Umaidah, Y., Adam, R. I., & Adam, R. I. (2024). Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3), 2755–2762. https://doi.org/10.23960/jitet.v12i3.4902

Triningsih, E. (2025). *ANALISIS SENTIMEN TERHADAP PROGRAM MAKAN BERGIZI GRATIS MENGGUNAKAN ALGORITMA MACHINE LEARNING PADA SOSIAL MEDIA X* [Universitas Islam Negeri Sultan Syarif Kasim Riau.]. https://doi.org/10.47065/bits.v9i9.999

Vebrian, Y. Z., & Kustiyono. (2025). A SENTIMENT ANALYSIS OF FREE MEAL PLANS ON SOCIAL MEDIA USING NAÏVE BAYES ALGORITHMS. *JURNAL INOVTEK POLBENG*, *10*(1), 355–366.

Yuspita, E., & Suryono, R. R. (2023). Perbandingan Berbagai Metode Klasifikasi Teks Untuk Sentimen Kebijakan Makan Gratis di Indonesia. *Indonesian Journal of Computer Science*, *12*(2), 284–301. https://doi.org/https://doi.org/10.33022/ijcs.v13i5.4440