



ANALISIS SENTIMEN MASYARAKAT TERHADAP PERMASALAHAN KERACUNAN PROGRAM MAKAN BERGIZI GRATIS (MBG) PADA SOSIAL MEDIA ‘X’

**Meita Dwi Anggraeni¹⁾, Alfian Chandra Yudiananta²⁾, Azizah Hasna Arifin³⁾
Wafi Arifin⁴⁾, Irham Mahfudz⁵⁾, Yanuar Tri Fauzi⁶⁾**

¹⁾ “Program Studi Sistem Informasi” STMIK BINA PATRIA Magelang

²⁾ “Program Studi Manajemen Informatika” STMIK BINA PATRIA Magelang

³⁾ “Program Studi Manajemen Informatika” STMIK BINA PATRIA Magelang

⁴⁾ “Program Studi Manajemen Informatika” STMIK BINA PATRIA Magelang

⁵⁾ “Program Studi Teknik Informatika” STMIK BINA PATRIA Magelang

⁶⁾ “Program Studi Teknik Informatika” STMIK BINA PATRIA Magelang

Email : meitadww@gmail.com¹⁾, alfianchandray@gmail.com²⁾, azizahhasna5@gmail.com³⁾
wafiarifin0609@gmail.com⁴⁾, yanuartrif@gmail.com⁵⁾, irhammahfudz01@gmail.com⁶⁾

Abstract

The Free Nutritious Meal Program introduced by President-elect Prabowo Subianto has garnered public attention and sparked a variety of responses, both positive and negative. The Free Nutritious Meal Program (MBG) is part of the Indonesian government's policy aimed at improving public nutrition, the welfare of school-going children, and the quality of human resources from an early age. However, during its implementation, several problems have emerged, including cases of food poisoning in several areas. Understanding public perception of these problems is important in order to evaluate the program's success. This study aims to analyze public sentiment towards the MBG program on social media "X" using three algorithms: SVM, K-Nearest Neighbor, and Naive Bayes. Data obtained through the crawling process is then processed with a preprocessing stage and sentiment classification into two categories: positive and negative. The evaluation results show that the Naive Bayes model has a higher accuracy than the other models, namely 86.70% accuracy, 86.10% precision, 86.70% recall, and 83.70% F1 score. The analysis shows that the number of negative sentiments is greater than positive sentiments. The results of this study can be used by the government to improve the socialization and implementation of programs to ensure greater public acceptance, and to emphasize the importance of implementing strict food safety standards, from raw material procurement to processing.

Keywords : *Data Mining, Sentiment Analysis, SVM, K-NN, Naive Bayes*

Abstrak

Program Makan Bergizi Gratis (MBG) menjadi sorotan publik dan memicu berbagai tanggapan, baik positif maupun negatif, seiring dengan pelaksanaannya saat ini. Program Makan Bergizi Gratis (MBG) menjadi bagian dari kebijakan pemerintah Indonesia yang bertujuan meningkatkan kondisi gizi masyarakat, kesejahteraan anak-anak yang sekolah, dan kualitas sumber daya manusia sejak usia dini. Namun, dalam pelaksanaannya, muncul beberapa permasalahan, termasuk kasus keracunan makanan di beberapa daerah. Persepsi masyarakat tentang adanya masalah ini penting untuk dipahami agar bisa mengevaluasi keberhasilan program tersebut. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap program MBG di media sosial “X” menggunakan tiga algoritma, yakni SVM, K-Nearest Neighbor, dan Naive Bayes. Data yang diperoleh melalui proses crawling kemudian diproses dengan tahap preprocessing dan klasifikasi sentimen ke dalam dua kategori, yaitu positif dan negatif. Hasil evaluasi menunjukkan



bahwa model Naive Bayes memiliki akurasi lebih tinggi dibandingkan model lainnya, yaitu 86.70% akurasi, 86.10% presisi, 86.70% recall, dan 83.70% F1 score. Analisis menunjukkan bahwa persentase atau jumlah sentimen negatif lebih banyak dibandingkan dengan sentimen positif. Hasil penelitian ini bisa digunakan pemerintah untuk memperbaiki sosialisasi dan implementasi program agar lebih diterima oleh masyarakat, serta menegaskan pentingnya penerapan standar keamanan pangan yang ketat mulai dari pengadaan bahan baku hingga proses pengolahan.

Kata kunci : Data Mining, Analisis Sentimen , SVM, K-NN, Naive Bayes

1. Pendahuluan

Program Makan Bergizi Gratis (MBG) adalah salah satu kebijakan penting dari pemerintah Indonesia yang bertujuan meningkatkan kesehatan masyarakat, kesejahteraan siswa, dan kualitas SDM sejak masa sekolah. Dengan memberikan makanan bergizi secara gratis, pemerintah berupaya mengatasi berbagai masalah gizi seperti stunting, anemia, dan kekurangan energi kronis yang masih dialami anak-anak di berbagai wilayah. Namun, dalam pelaksanaan program ini, muncul beberapa masalah yang menimbulkan kekhawatiran masyarakat, terutama terkait kasus keracunan makanan di beberapa daerah. Beberapa peristiwa di berbagai tempat menunjukkan adanya kelemahan dalam pengawasan kualitas, distribusi, dan kebersihan makanan. Kasus ini bukan hanya berdampak pada kesehatan para korban, namun juga mengurangi kepercayaan masyarakat terhadap keberhasilan program MBG. Masalah keracunan ini menunjukkan betapa pentingnya penerapan standar keamanan pangan yang ketat, mulai dari pengadaan bahan baku, pengolahan, hingga pendistribusian makanan kepada anak-anak sekolah. Selain itu, transparansi, akuntabilitas, dan keterlibatan pihak-pihak terkait seperti pemerintah daerah, penyedia jasa katering, serta lembaga pengawas makanan menjadi faktor penting untuk menjamin keberhasilan dan kelangsungan program MBG. Kasus ini kemudian menjadi perhatian utama di media sosial, termasuk di platform “X” yang merupakan wadah opini publik yang aktif dan dinamis.

Melalui platform “X”, masyarakat bisa menyampaikan berbagai tanggapan, mulai dari dukungan terhadap program makan bergizi gratis hingga kritik tajam terhadap cara pengelolaan, kebersihan, dan pengawasan dalam pelaksanaannya. Fenomena ini menunjukkan bahwa pendapat yang muncul di dunia maya bisa menjadi tolak ukur penting dalam menilai seberapa baik suatu kebijakan sosial berjalan. Namun, pendapat yang beredar di media sosial sering kali bersifat tidak lengkap, penuh emosi, dan sulit dinilai secara objektif. Oleh karena itu, dibutuhkan pendekatan berbasis data yang mampu mengolah dan menganalisis teks pendapat masyarakat secara terstruktur. Salah satu metode yang bisa digunakan adalah analisis sentimen, yaitu teknik dalam bidang pemrosesan bahasa alami yang digunakan untuk mengenali perasaan atau sikap seseorang dalam teks, apakah positif atau negatif.

Dengan menganalisis sentimen masyarakat dari unggahan dan komentar di platform “X” terkait isu keracunan dalam program MBG, peneliti bisa memahami secara lebih jelas bagaimana masyarakat memandang isu tersebut, serta mengetahui topik utama yang memengaruhi kepercayaan mereka terhadap kebijakan pemerintah.

Penelitian ini bertujuan untuk menggunakan tiga algoritma yaitu SVM, K-Nearest Neighbor, dan Naive Bayes dalam menganalisis sentimen masyarakat di media sosial X mengenai masalah keracunan yang terjadi dalam program makan siang gratis. Penelitian ini akan menggunakan data dari media sosial X untuk mengenali dan mengelompokkan pendapat masyarakat mengenai masalah tersebut. Dengan adanya hasil penelitian ini, diharapkan dapat memberikan pemahaman yang lebih baik mengenai tanggapan



masyarakat terhadap masalah yang sedang terjadi. Selain itu, penelitian ini juga akan menghasilkan peta sentimen masyarakat terhadap Program MBG, yang bisa menjadi bahan evaluasi bagi pemerintah dalam meningkatkan kualitas, transparansi, dan keamanan program makan bergizi gratis di masa depan. Hasil penelitian ini juga diharapkan dapat menjadi pertimbangan dalam merancang kebijakan yang lebih baik dan efektif serta lebih disambut oleh masyarakat.

2. Kajian Literatur

2.1 Analisis Sentimen

Analisis sentimen adalah bagian dari text mining yang bertujuan untuk mengenali dan mengelompokkan pendapat atau perasaan yang terdapat dalam teks, seperti ulasan produk, komentar di media sosial, atau artikel berita. Menurut Bing Liu (2015), analisis sentimen adalah bidang yang mempelajari pendapat, perasaan, penilaian, sikap, dan emosi seseorang terhadap sesuatu atau sifat-sifatnya. Pendapat yang sama juga disampaikan oleh Medhat, Hassan, dan Korashy (2014) yang menyatakan bahwa analisis sentimen bertujuan untuk mengetahui sikap penulis atau pembicara terhadap suatu topik dengan melihat keseluruhan teks, apakah bersifat positif atau negatif. Feldman (2013) menjelaskan bahwa analisis sentimen merupakan proses mengenali, mengekstrak, serta mengklasifikasikan pendapat, perasaan, dan sikap yang terdapat dalam teks. Dalam konteks penelitian berbahasa Indonesia, Septiadi dan Sari (2021) menyebut analisis sentimen sebagai proses pengolahan teks untuk mengenali kecenderungan opini publik terhadap sesuatu, baik positif maupun negatif. Dengan demikian, analisis sentimen dapat diartikan sebagai pendekatan komputasi untuk menilai sikap atau opini seseorang terhadap topik tertentu berdasarkan teks, yang melibatkan beberapa tahap seperti preprocessing, ekstraksi fitur, dan klasifikasi menggunakan algoritma pembelajaran mesin seperti Support Vector Machine (SVM), Naive Bayes, atau K-Nearest Neighbors (KNN).

2.2 Algoritma SVM

Algoritma Support Vector Machine (SVM) adalah salah satu metode pembelajaran terarah yang sering digunakan dalam analisis sentimen karena kemampuannya dalam memproses data yang memiliki banyak dimensi dan mampu memberikan hasil yang akurat. Cara kerja algoritma SVM adalah dengan mencari sebuah garis pemisah optimal yang dapat membagi data dari dua kategori berbeda dengan jarak yang paling lebar (Cortes & Vapnik, 1995). Dalam analisis sentimen berbasis teks, SVM digunakan untuk mengelompokkan pendapat menjadi dua kategori, yaitu positif atau negatif, berdasarkan fitur yang digunakan seperti bag-of-words atau TF-IDF (Pang, Lee, & Vaithyanathan, 2002).

2.3 Algoritma K-NN

Algoritma K-Nearest Neighbor (K-NN) adalah metode pembelajaran terawasi yang digunakan untuk klasifikasi dan regresi. Prinsip utama algoritma ini adalah menentukan kelas dari data baru berdasarkan tingkat kesamaan jarak dengan beberapa data terdekat (k tetangga) dalam dataset pelatihan (Cover & Hart, 1967). Menurut Han, Kamber, dan Pei (2012), KNN termasuk dalam kategori pembelajaran lambat (lazy learning) karena tidak membuat model selama proses pelatihan, melainkan menyimpan semua data pelatihan dan melakukan klasifikasi saat data uji diberikan. Algoritma ini mengklasifikasikan data uji berdasarkan kelas mayoritas dari k data pelatihan terdekat, yang dihitung dengan menggunakan jarak seperti Euclidean Distance (Tan, Steinbach, &



Kumar, 2019). Dalam analisis sentimen, K-NN digunakan untuk mengklasifikasikan teks ke dalam kategori positif, negatif, atau netral berdasarkan kesamaan vektor fitur teks baru dengan data pelatihan. Zhang, Wang, dan Liu (2011) menyatakan bahwa K-NN cukup efektif untuk data teks berdimensi tinggi seperti opini atau ulasan, terutama jika

representasi datanya menggunakan TF-IDF atau bag-of-words. Meskipun sederhana, algoritma K-NN memiliki beberapa kelebihan seperti tidak memerlukan asumsi distribusi data dan dapat berfungsi baik untuk data yang tidak linear. Namun, K-NN juga memiliki kelemahan, seperti kompleksitas komputasi yang tinggi pada dataset besar serta sensitivitas terhadap pemilihan nilai k dan skala fitur (Han et al., 2012).

2.4 Algoritma Naive Bayes

Algoritma Naive Bayes adalah salah satu metode pembelajaran mesin yang menggunakan pendekatan probabilistik berdasarkan Teorema Bayes. Metode ini mengasumsikan bahwa setiap fitur dalam data tidak saling bergantung satu sama lain. Menurut Han, Kamber, dan Pei (2012), Naive Bayes adalah pengklasifikasi statistik yang berlandaskan asumsi bahwa setiap atribut mempengaruhi kelas tertentu secara terpisah. Dengan kata lain, algoritma ini memprediksi kelas suatu data dengan memperhitungkan kemungkinan munculnya fitur dalam data terhadap kelas tertentu. Mitchell (1997) menjelaskan bahwa algoritma Naive Bayes menghitung kemungkinan suatu kelas berdasarkan data masukan dengan menggunakan aturan Bayes, di mana kemungkinan kelas meningkat jika suatu fitur sering muncul pada kelas tersebut. Meskipun asumsi independensi fitur jarang terpenuhi secara sempurna, Zhang (2004) menegaskan bahwa algoritma Naive Bayes tetap menunjukkan kinerja yang baik dalam berbagai kasus klasifikasi, terutama dalam klasifikasi teks seperti analisis sentimen. Dalam konteks analisis sentimen, Naive Bayes digunakan untuk mengklasifikasikan opini atau ulasan menjadi kategori positif, negatif, atau netral berdasarkan distribusi kata dalam teks. Pang, Lee, dan Vaithyanathan (2002) menunjukkan bahwa Naive Bayes adalah salah satu algoritma pertama yang digunakan untuk menganalisis sentimen ulasan film dengan hasil akurasi yang cukup baik dibandingkan dengan algoritma lain seperti SVM. Keunggulan utama algoritma ini adalah kecepatan proses pelatihan, mudah diimplementasikan, serta mampu menangani data berdimensi tinggi seperti teks ulasan atau komentar di media sosial.

2.5 Program Makan Bergizi Gratis (MBG)

Program Makan Bergizi Gratis (MBG) adalah salah satu prioritas pemerintah Indonesia yang bertujuan meningkatkan kesehatan gizi anak-anak sekolah dan memperkuat ketahanan pangan masyarakat. Program ini memberikan makanan bergizi secara gratis kepada siswa, agar bisa tumbuh menjadi generasi yang sehat, cerdas, dan berkualitas. Menurut Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan, MBG adalah upaya pemerintah untuk meningkatkan kualitas sumber daya manusia dengan memberi makanan bergizi kepada anak sekolah dan masyarakat berpenghasilan rendah, sebagai cara mengatasi masalah stunting, anemia, dan kekurangan gizi. Selaras dengan hal itu, program MBG juga menjadi kebijakan strategis pemerintah yang tidak hanya meningkatkan gizi anak, tetapi juga mendorong pemberdayaan ekonomi daerah melalui kerja sama dengan pengusaha makanan lokal, sehingga membantu meningkatkan ketahanan pangan nasional. Kementerian Kesehatan Republik Indonesia (2024) menyatakan bahwa pelaksanaan Program MBG harus didasarkan pada prinsip gizi seimbang, yaitu mencakup kebutuhan energi, protein, vitamin, dan mineral sesuai dengan

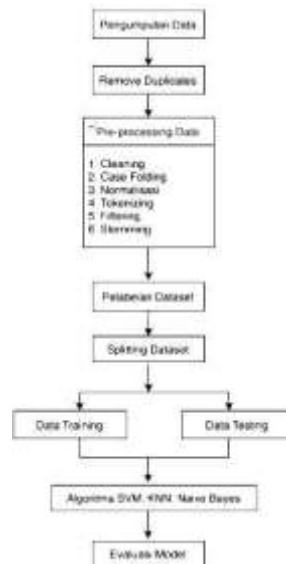
usia anak, serta memperhatikan kebersihan dan keamanan pangan di setiap tahap penyediannya.

2.6 Studi Terkait

Penelitian mengenai analisis sentimen sudah banyak dilakukan dengan berbagai algoritma klasifikasi, seperti Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), dan Naive Bayes (NB). Ketiga algoritma ini masing-masing memiliki keunggulan dalam mengklasifikasikan opini atau sentimen dari data teks, termasuk dalam konteks program makan bergizi gratis (MBG) dan kasus keracunan makanan. Fatkhirrohman (2025) dalam penelitiannya yang berjudul "Analisis Sentimen Program Makan Bergizi Gratis Pemerintah" menggunakan algoritma SVM. Penelitian ini memanfaatkan data dari media sosial untuk mengetahui persepsi masyarakat terhadap program MBG. Hasilnya menunjukkan kebanyakan masyarakat memiliki sentimen positif, meski ada juga opini negatif yang berkaitan dengan masalah higienitas makanan dan risiko keracunan. Selain itu, penelitian oleh Pratama et al. (2024) dengan judul "Analisis Sentimen Program Makan Siang dan Susu Gratis Menggunakan Naive Bayes" menggunakan algoritma Naive Bayes untuk mengklasifikasikan opini publik mengenai kebijakan makan gratis bagi siswa. Model NB berhasil menghasilkan akurasi yang tinggi dengan proses yang cepat, meski kurang baik dalam memahami konteks dari bahasa informal di media sosial. Sementara itu, penelitian oleh Rahmawati dan Nugroho (2023) terhadap ulasan restoran daring menunjukkan bahwa algoritma SVM memiliki akurasi tertinggi dibandingkan KNN dan Naive Bayes pada dataset berbasis TF-IDF. Penelitian ini menyimpulkan bahwa SVM lebih efektif dalam menangani data berdimensi tinggi dan sentimen yang kompleks.

3. Metode Penelitian

Gambar berikut adalah ilustrasi yang memperlihatkan tahapan penelitian yang dilakukan oleh peneliti.



Gambar 1. Tahapan Alur Penelitian

Penelitian ini dimulai dengan pengumpulan data, dimana data diperoleh melalui proses crawling menggunakan tweepy di Google Colab dengan Python. Python adalah bahasa pemrograman yang mudah dipelajari dan bisa dijalankan di berbagai platform,



dengan fokus pada kejelasan dan kemudahan dalam membaca kode. Setelah itu, data yang terduplikat dihilangkan. Data tersebut kemudian menjadi dataset yang digunakan dalam penelitian. Dataset tersebut harus melewati tahap preprocessing untuk menghapus data yang tidak relevan. Data harus diolah agar bisa digunakan. Setelah proses preprocessing selesai, dataset siap diimplementasikan dengan beberapa model untuk analisis sentimen. Evaluasi dilakukan pada setiap model untuk mengetahui seberapa akurat model tersebut dalam menganalisis sentimen.

Proses pengumpulan data dilakukan dalam beberapa tahap. Setelah data berhasil dikumpulkan, langkah selanjutnya adalah filtering untuk menghilangkan tweet yang tidak relevan, seperti spam atau iklan, agar hanya opini yang benar-benar membahas masalah keracunan pada program makan bergizi gratis yang digunakan. Setelah itu, data menjalani tahap preprocessing yang mencakup penghapusan karakter khusus, angka, tanda baca, kata yang tidak penting, URL, dan emoji. Selain itu, dilakukan proses normalisasi, tokenizing, dan stemming untuk mengubah teks menjadi bentuk dasar kata, sehingga meningkatkan akurasi dalam analisis sentimen. Tahapan analisis data dalam penelitian ini dimulai dengan preprocessing untuk membersihkan data dan memastikan hanya kata-kata yang memiliki makna yang digunakan dalam analisis. Setelah itu, dilakukan klasifikasi sentimen menggunakan algoritma SVM, KNN, dan Naive Bayes, di mana model ini dilatih menggunakan dataset latih sebelum diterapkan pada data yang telah dikumpulkan. Tujuan dari proses klasifikasi adalah mengelompokkan opini publik ke dalam dua kategori, yaitu positif dan negatif. Setelah selesai melakukan klasifikasi, hasil analisis sentimen disajikan dalam beberapa bentuk, seperti diagram batang untuk menunjukkan distribusi sentimen masyarakat, serta tabel yang berisi contoh tweet dari setiap kategori sentimen. Selain itu, dilakukan interpretasi terhadap hasil untuk mengidentifikasi faktor-faktor yang memengaruhi sentimen masyarakat terhadap masalah keracunan pada program makan bergizi gratis, agar bisa dipakai sebagai bahan evaluasi kebijakan pemerintah.

4. Hasil dan Pembahasan

4.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini didapatkan melalui proses crawling menggunakan tweepy di GoogleColab. Dalam proses pengambilan data, dilakukan pencarian terhadap tweet yang membicarakan isu keracunan pada program makan siang gratis. Sumber data berasal dari media sosial “X”, dengan kata kunci yang digunakan adalah “keracunan program makan bergizi gratis”. Hasil crawling berupa 1016 tweet yang diperoleh dalam rentang waktu September hingga Oktober 2025. Data tersebut mencakup berbagai pendapat dari masyarakat mengenai permasalahan keracunan di program tersebut, baik yang bersifat positif maupun negatif. Selanjutnya, data yang telah dikumpulkan disimpan dalam bentuk file .csv yang memiliki beberapa kolom, seperti created_at, id_str, full_text, quote_count, reply_count, retweet_count, favorite_count, lang, user_id_str, conversation_id_str, tweet_url, dan username”. Namun, dalam penelitian ini hanya kolom “full_text” yang digunakan untuk analisis sentimen karena kolom tersebut berisi kalimat lengkap dari setiap tweet yang akan dianalisis.

4.2 Preprocessing Data

Sebelum data dianalisis, beberapa langkah preprocessing dilakukan untuk membersihkan dan mempersiapkan teks agar algoritma Naive Bayes bisa lebih mudah diproses. Langkah-langkahnya adalah sebagai berikut:

- a. Proses cleaning: Proses pembersihan data tweet dilakukan dengan menghilangkan

atribut yang tidak perlu, seperti mention, URL, hashtag, dan *emoticon*.

Tabel. 1. Proses Cleaning

Sebelum	Sesudah
Jaringan Pemantau Pendidikan Indonesia (JPPI) bersama para relawan terus melakukan pemantauan dan menerima laporan dari berbagai daerah terkait program Makan Bergizi Gratis (MBG). https://t.co/GQxuFaN2eJ	Jaringan Pemantau Pendidikan Indonesia JPPI bersama para relawan terus melakukan pemantauan dan menerima laporan dari berbagai daerah terkait program Makan Bergizi Gratis MBG

- a. Case Folding adalah proses mengubah semua teks menjadi huruf kecil agar data tetap konsisten dan tidak terjadi perbedaan makna karena perbedaan antara huruf besar dan kecil.

Tabel 2. Proses Case Folding

Sebelum	Sesudah
GERAK CEPAT ATASI KERACUNAN . Program Makan Bergizi Gratis (MBG) terus dijalankan dengan sigap dan terukur. Seperti pesan Presiden Prabowo Subianto MBG bukan sekadar bagi-bagi makanan tapi langkah strategis mencetak Generasi Emas Indonesia 2045. https://t.co/zk58C2CN0q	gerak cepat atasi keracunan program makan bergizi gratis mbg terus dijalankan dengan sigap dan terukur seperti pesan presiden prabowo subianto mbg bukan sekadar bagibagi makanan tapi langkah strategis mencetak generasi emas indonesia 2045

- b. Normalisasi adalah proses mengubah kata-kata yang tidak baku atau singkatan menjadi ke dalam bentuk kata yang baku, misalnya seperti "yg" menjadi "yang" dan "thn" menjadi "tahun".

Tabel 3. Proses Normalization

Sebelum	Sesudah
Bayangin aja deehhhh banyak anak dan ibu hamil yg jadi lebih sehat krn Program Makan Bergizi Gratis ini . Capaian1thn PemerintahanPrabowo https://t.co/3WqH8G7rTR	bayangin aja deehhhh banyak anak dan ibu hamil yang jadi lebih sehat karena program makan bergizi gratis ini capaian1tahun pemerintahanprabowo

- c. Tokenisasi adalah proses memecah teks menjadi kata-kata atau unit-unit kecil yang disebut token, sehingga teks tersebut lebih mudah untuk dianalisis.

Tabel 4. Proses Tokenize

Sebelum	Sesudah
jaringan pemantau pendidikan indonesia jppi bersama para relawan terus melakukan pemantauan dan menerima laporan dari berbagai daerah terkait program makan bergizi gratis mbg	['jaringan', 'pemantau', 'pendidikan', 'indonesia', 'jppi', 'bersama', 'para', 'relawan', 'terus', 'melakukan', 'pemantauan', 'dan', 'menerima', 'laporan', 'dari', 'berbagai', 'daerah', 'terkait', 'program', 'makan', 'bergizi', 'gratis', 'mbg']

d. Hapus Stopword adalah menghilangkan kata-kata biasa (stopwords) seperti "dan", "dari", dan lainnya yang tidak memiliki arti penting dalam analisis sentimen.

Tabel 5. Proses Stopword

Sebelum	Sesudah
'jaringan', 'pemantau', 'pendidikan', 'indonesia', 'jppi', 'bersama', 'para', 'relawan', 'terus', 'melakukan', 'pemantauan', 'dan', 'menerima', 'laporan', 'dari', 'berbagai', 'daerah', 'terkait', 'program', 'makan', 'bergizi', 'gratis', 'mbg'	'jaringan', 'pemantau', 'pendidikan', 'indonesia', 'jppi', 'relawan', 'pemantauan', 'menerima', 'laporan', 'daerah', 'terkait', 'program', 'makan', 'bergizi', 'gratis', 'mbg'

e. Stemming adalah proses mengubah kata-kata menjadi bentuk dasar atau bentuk akar katanya, seperti contoh "laporan" diubah menjadi "lapor" sehingga analisis menjadi lebih akurat.

Tabel 5. Proses Stemming

Sebelum	Sesudah
'jaringan', 'pemantau', 'pendidikan', 'indonesia', 'jppi', 'relawan', 'pemantauan', 'menerima', 'laporan', 'daerah', 'terkait', 'program', 'makan', 'bergizi', 'gratis', 'mbg'	jaring pantau didik indonesia jppi rawan pantau terima lapor daerah kait program makan gizi gratis mbg

4.3. Labeling Sentimen

Setelah proses persiapan data selesai, data kemudian diberi label sentimen untuk mengetahui jenis opini pada setiap tweet. Proses memberi label dimulai dengan menentukan apakah setiap tweet memiliki sentimen positif atau negatif terhadap program tersebut. Sentimen positif menunjukkan dukungan atau pandangan yang baik terhadap program makanan bergizi gratis, sedangkan sentimen negatif menunjukkan kekhawatiran atau kritik akibat adanya masalah keracunan pada program tersebut. Peneliti menggunakan library IndoBERT digunakan untuk membantu labeling data latih, bukan sebagai bukan untuk klasifikasi akhir model yang dibandingkan. Setelah itu, peneliti memeriksa kembali hasil klasifikasi untuk memastikan keakuratannya. Jika ada kesalahan dalam klasifikasi, peneliti melakukan perbaikan secara manual agar data lebih tepat. Selanjutnya, data dibagi menjadi dua bagian, yaitu data latih dan data uji dengan perbandingan 80:20, untuk digunakan dalam melatih dan menguji model analisis sentimen.

Tabel 6. Labeling Data

Tweet	Label
program makan bergizi gratis mbg menjadi sorotan masyarakat sejak diluncurkan pada januari 2025 silam hingga kini menuai sejumlah masalah kasus utamanya adalah ribuan penerima manfaat yakni para siswa bahkan guru di beberapa daerah keracunan setelah menyantap makanan	Negatif
presiden prabowo subianto langsung bertindak tanggap dan gercep begitu menerima laporan kasus keracunan terkait program makan bergizi gratis mbg beliau memerintahkan investigasi menyeluruh memastikan korban mendapat perawatan optimal dan evaluasi standar dapur mbg	Positif

Hasil pelabelan ini kemudian digunakan sebagai data testing/latih untuk proses pembuatan model analisis sentimen.

4.3. Pembagian Data Latih dan Data Uji

Pembagian data ini dilakukan dengan rasio tertentu, seperti 80% untuk data latih dan 20% untuk data uji. Dengan demikian, model dapat belajar dari sebagian besar data yang tersedia dan mengujinya dengan data yang belum pernah dilihat sebelumnya, sehingga dapat secara objektif mengevaluasi akurasi dan kinerja model sebelum menerapkannya pada data baru.

```

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB

X = cleaned_data['text'].values
y = cleaned_data['label']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)

print("Training data shape: ", X_train.shape)
print("Training data size: ", len(X_train))
print("Testing data shape: ", X_test.shape)
print("Testing data size: ", len(X_test))

vectorizer = CountVectorizer()
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)

X_train_vec.shape, X_test_vec.shape

# Output: (1000, 1000), (200, 1000)

# (1000, 2000), (200, 2000)

```

Gambar 1. Program Model SVM, KNN, Naive Bayes

4.3. Model SVM, K-NN dan Naive Bayes

Proses pelatihan model SVM, K-NN, dan Naive Bayes melalui beberapa tahap. Tahap pertama adalah persiapan data, yaitu mengubah teks mentah menjadi bentuk numerik menggunakan teknik CountVectorizer. Langkah ini bertujuan untuk mengubah sekumpulan dokumen teks menjadi matriks angka dengan menghitung berapa kali setiap kata muncul dalam setiap dokumen. Ini adalah langkah penting dalam pemrosesan bahasa alami (NLP) agar data teks dapat dibaca dan diproses oleh mesin. Tahap kedua adalah membagi data menjadi dua bagian, yaitu data latih dan data uji. Data uji digunakan untuk membangun model, sedangkan data latih digunakan untuk mengevaluasi kinerja model tersebut.

Setelah model selesai dilatih, langkah berikutnya adalah melakukan prediksi sentimen terhadap data uji. Model yang sudah dipelajari akan menganalisis karakteristik dari data tersebut dan membagi setiap teks ke dalam kategori sentimen yang sudah ditentukan sebelumnya, seperti positif atau negatif. Hasil prediksi ini kemudian dinilai menggunakan beberapa metrik seperti akurasi, precision, recall, dan F1 Score untuk mengetahui seberapa baik model dalam membedakan sentimen. Dengan cara ini, sistem bisa secara otomatis mengelompokkan teks berdasarkan sentimen yang terdapat di dalamnya, sehingga memudahkan proses analisis dalam berbagai bidang yang menggunakan teknologi NLP.

```

from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.ensemble import RandomForestClassifier
import numpy as np
import matplotlib.pyplot as plt

models = [
    ("SVC", SVC(kernel="linear", random_state=42)),
    ("KNN", KNeighborsClassifier(n_neighbors=4)),
    ("RF", RandomForestClassifier(n_estimators=100))
]

results = []
for model_name, model in models:
    print(f"Model: {model_name} training...")
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test, y_val)
    results.append((model_name, y_pred))
    print(f"Model: {model_name} accuracy: {accuracy_score(y_val, y_pred)}\n")
    print(f"Model: {model_name} classification report: {classification_report(y_val, y_pred)}\n")
    print(f"Model: {model_name} confusion matrix: {confusion_matrix(y_val, y_pred)}\n")

```

Gambar 2. Model SVM, KNN, Naive Bayes

4.4. Evaluasi

Setelah prediksi selesai, hasil dari prediksi sentimen yang digunakan algoritma SVM, KNN, dan Naive Bayes dinilai. Penilaian ini dilakukan dengan menggunakan beberapa metrik seperti akurasi, presisi, recall, dan F1 Score. Hasil penilaian tersebut menunjukkan seberapa baik masing-masing model dalam mengklasifikasikan sentimen secara tepat.

Classification Report for SVM:				
	precision	recall	f1-score	support
NEGATIVE	0.892	0.929	0.910	169.000
POSITIVE	0.556	0.441	0.492	34.000
accuracy	0.847	0.847	0.847	0.847
macro avg	0.724	0.685	0.701	203.000
weighted avg	0.836	0.847	0.840	203.000

Gambar 3. Hasil SVM

Model SVM memiliki akurasi 84,70%, presisi 83,60%, recall 84,70%, dan F1-Score 84,00%. Hasil ini menunjukkan bahwa model tersebut cukup efektif dalam mengklasifikasikan sentimen.

Classification Report for KNN:				
	precision	recall	f1-score	support
NEGATIVE	0.864	0.976	0.917	169.000
POSITIVE	0.667	0.235	0.348	34.000
accuracy	0.852	0.862	0.852	0.852
macro avg	0.765	0.606	0.632	203.000
weighted avg	0.831	0.862	0.821	203.000

Gambar 4. Hasil KNN

Model K-NN memiliki akurasi 85,20%, presisi 83,10%, recall 85,20%, serta F1-Score 82,10%. Hasil ini menunjukkan bahwa model tersebut cukup efektif dalam mengklasifikasikan sentimen.

Classification Report for Naive Bayes:				
	precision	recall	f1-score	support
NEGATIVE	0.870	0.988	0.925	169.000
POSITIVE	0.818	0.265	0.400	34.000
accuracy	0.887	0.867	0.887	0.887
macro avg	0.844	0.626	0.663	203.000
weighted avg	0.861	0.867	0.837	203.000

Gambar 3. Hasil Naive Bayes

Model Naive Bayes memiliki akurasi 86,70%, presisi 86,10%, recall 86,70%, dan F1-Score 83,70%. Hasil ini menunjukkan bahwa model tersebut cukup baik dalam mengklasifikasikan sentimen dibandingkan dengan model SVM dan K-NN.



5. Kesimpulan

Hasil dari analisis sentimen yang dilakukan terhadap masalah keracunan dalam program makan bergizi gratis di media sosial “X” menggunakan tiga algoritma yaitu SVM, K-NN, dan Naive Bayes menunjukkan bahwa model Naive Bayes memiliki performa terbaik dengan akurasi 86.70%, presisi 86.10%, recall 86.70%, dan F1-Score 83.70%. Hasil klasifikasi menunjukkan bahwa model lebih mampu mengenali sentimen negatif dibandingkan sentimen positif. Nilai recall tertinggi terdapat pada kategori negatif (0.988), sedangkan nilai recall terendah ada pada kategori positif (0.265). Dari hasil analisis tersebut, dapat disimpulkan bahwa opini masyarakat di media sosial X mengenai masalah keracunan dalam program makan bergizi gratis lebih banyak bersifat negatif. Kehawatiran masyarakat terhadap penerapan standar keamanan pangan yang ketat dalam proses pengadaan bahan baku dan pengolahan perlu diperhatikan. Dengan demikian, dapat disimpulkan bahwa algoritma SVM, K-NN, dan Naive Bayes merupakan pendekatan yang efektif dalam menganalisis sentimen publik, terutama dalam konteks kesehatan dan pangan. Dalam kasus keracunan makanan di program MBG, pendekatan ini berpotensi membantu pemerintah atau instansi terkait dalam mendeteksi keluhan sejak awal, mengevaluasi kepuasan publik, serta meningkatkan kualitas pelaksanaan program.

Daftar Pustaka

- Anwar, S., & Rahmawati, I. (2022). Analisis Sentimen terhadap Kebijakan Publik Menggunakan Metode Naïve Bayes Classifier pada Media Sosial Twitter. *Jurnal Teknologi Informasi dan Komunikasi*, 14(2), 101–112.
- Budiarto, D., & Lestari, W. (2021). Penerapan Analisis Sentimen Menggunakan Lexicon Based untuk Menilai Opini Publik pada Program Pemerintah. *Jurnal Ilmiah Informatika dan Komputer*, 9(1), 55–64.
- Departemen Kesehatan Republik Indonesia. (2024). Pedoman Pelaksanaan Program Makan Bergizi Gratis (MBG) untuk Peserta Didik Sekolah Dasar dan Menengah Pertama. Jakarta: Kemenkes RI.
- Fitriyah, N., & Rakhman, A. (2023). Pemanfaatan Analisis Sentimen untuk Evaluasi Program Sosial Pemerintah di Media Sosial Twitter. *Jurnal Transformasi Digital*, 2(1), 43–57.
- Mayasari, I., & Nugroho, A. (2022). Implementasi Natural Language Processing (NLP) untuk Analisis Sentimen Opini Publik terhadap Isu Sosial. *Jurnal Informatika dan Aplikasi Komputer*, 12(3), 67–75.
- Nugraha, D., & Astuti, R. (2021). Perbandingan Metode Naïve Bayes dan SVM untuk Analisis Sentimen pada Opini Masyarakat. *Jurnal Sistem Cerdas*, 10(2), 90–98.
- Putri, M. D., & Sari, E. P. (2023). Evaluasi Persepsi Masyarakat terhadap Program Bantuan Pemerintah Menggunakan Analisis Sentimen dan Visualisasi Data. *Jurnal Informatika dan Data Science*, 4(1), 12–25.
- Satria, A., & Hidayat, T. (2022). Implementasi Text Mining dalam Analisis Sentimen Masyarakat terhadap Program Pemerintah Menggunakan Python. *Jurnal Komputer dan Sistem Informasi*, 20(3), 198–210.
- Supriyanto, D., & Dewi, R. (2023). Analisis Sentimen terhadap Program Bantuan Sosial Pemerintah Menggunakan Algoritma Machine Learning. *Jurnal Teknologi dan Aplikasi Komputer*, 11(1), 22–35.
- Wibisono, H., & Yulianto, A. (2022). Analisis Sentimen Masyarakat terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Teknologi Informasi dan Sains Data*, 3(2), 44–56.



- Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82–89. <https://doi.org/10.1145/2436256.2436274>
- Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions* (2nd ed.). 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. <https://doi.org/10.1016/j.asej.2014.04.011>
- Septiadi, A., & Sari, D. N. (2021). Analisis sentimen opini publik terhadap kebijakan pemerintah menggunakan algoritma Naïve Bayes. *Jurnal Informatika dan Teknologi*, 5(2), 45–52.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273– 297. <https://doi.org/10.1007/BF00994018>
- Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *Stanford University Technical Report*.
- Liem, S., Setiawan, T., & Pribadi, M. R. (2023). Comparison of SVM and Naïve Bayes Algorithms in Sentiment Analysis of TikTok Comments on Skincare Products. *AICOM Proceedings*.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing*, 79–86.
- Wibowo, A., & Putra, D. (2021). Analisis Sentimen pada Ulasan Produk Menggunakan Support Vector Machine. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(4), 713– 722.
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Tan, P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to data mining* (2nd ed.). Pearson.
- Zhang, L., Wang, S., & Liu, B. (2011). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(4), 1–15.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing*, 79–86.
- Zhang, H. (2004). The optimality of Naive Bayes. *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, 562–567.
- Fatkurrohman, A. (2025). Analisis Sentimen Program Makan Bergizi Gratis Pemerintah Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Informatika dan Sistem Cerdas*, 5(2), 45–55.
- Pratama, D., Sari, R., & Wulandari, F. (2024). Analisis Sentimen Program Makan Siang dan Susu Gratis Menggunakan Naive Bayes. *Jurnal Teknologi Informasi dan Komunikasi*, 12(1), 23–32.
- Rahmawati, N., & Nugroho, A. (2023). Perbandingan Algoritma SVM, KNN, dan Naive Bayes untuk Analisis Sentimen Ulasan Restoran Daring. *Jurnal Ilmiah Teknologi Informasi*, 7(3), 78–86.