

PENERAPAN METODE NAIVE BAYES DALAM SENTIMEN ANALISIS TERHADAP PROGRAM MAKAN SIANG DAN SUSU GRATIS DI MEDIA X

Jonathan De Agustino Wulla¹, Rino^{2*}

^{1,2} Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Buddhi Dharma

*Corresponding Author, email: rino@ubd.ac.id

ABSTRAK

Selama kampanye pemilihan presiden Indonesia tahun 2024, Program Makan Siang dan Susu Gratis menjadi salah satu isu kebijakan utama yang mencuri perhatian publik, terutama di media sosial X (sebelumnya Twitter). Banyaknya opini publik yang beredar menciptakan tantangan tersendiri bagi para pembuat kebijakan dalam memahami persepsi masyarakat secara objektif dan berbasis data. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap program tersebut dengan menerapkan algoritma Naïve Bayes dan metode pembobotan Term Frequency–Inverse Document Frequency (TF-IDF). Sebanyak 1.084 tweet berbahasa Indonesia dikumpulkan secara otomatis menggunakan pemrograman di Google Colab melalui pustaka tweet-harvest dan autentikasi token. Data yang diperoleh kemudian diproses melalui tahapan praproses teks, meliputi pembersihan karakter, normalisasi huruf, tokenisasi, penghapusan stopword, dan stemming untuk menyiapkan data agar dapat dianalisis secara efektif. Selanjutnya, data diubah menjadi representasi numerik menggunakan TF-IDF, lalu diklasifikasikan ke dalam dua kategori sentimen, yaitu positif dan negatif. Proses klasifikasi dilakukan menggunakan algoritma Naïve Bayes, dengan skema evaluasi menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa model mampu mencapai akurasi tinggi dalam mengidentifikasi arah sentimen publik. Temuan ini tidak hanya menggambarkan persepsi masyarakat terhadap program yang diusulkan, tetapi juga membuktikan efektivitas metode machine learning dalam menganalisis opini publik pada isu-isu sosial dan politik. Penelitian ini diharapkan dapat menjadi acuan dalam pengambilan keputusan berbasis data di masa mendatang.

Kata kunci: Analisis Sentimen, *Naïve Bayes*, Program Makan Siang, Susu Gratis.

I. PENDAHULUAN

Saat ini, teknologi berkembang pesat dan mempermudah akses informasi melalui berbagai platform media sosial. Sebagian besar informasi di media sosial berbentuk teks, berisi pujian, kritik, ujaran kebencian, hingga rumor. Survei APJII memprediksi pengguna internet Indonesia akan mencapai 221,5 juta pada 2024 (APJII, 2024)(Nursingah et al., 2024). Masyarakat kini dikenal sebagai netizen atau warganet, yang aktif menyampaikan opini dan perasaan mereka melalui komentar di media sosial (Saleh & Furbani, 2022). Pada Februari 2024, dalam Pemilu Presiden, pasangan calon nomor urut 02 mempromosikan program Makan

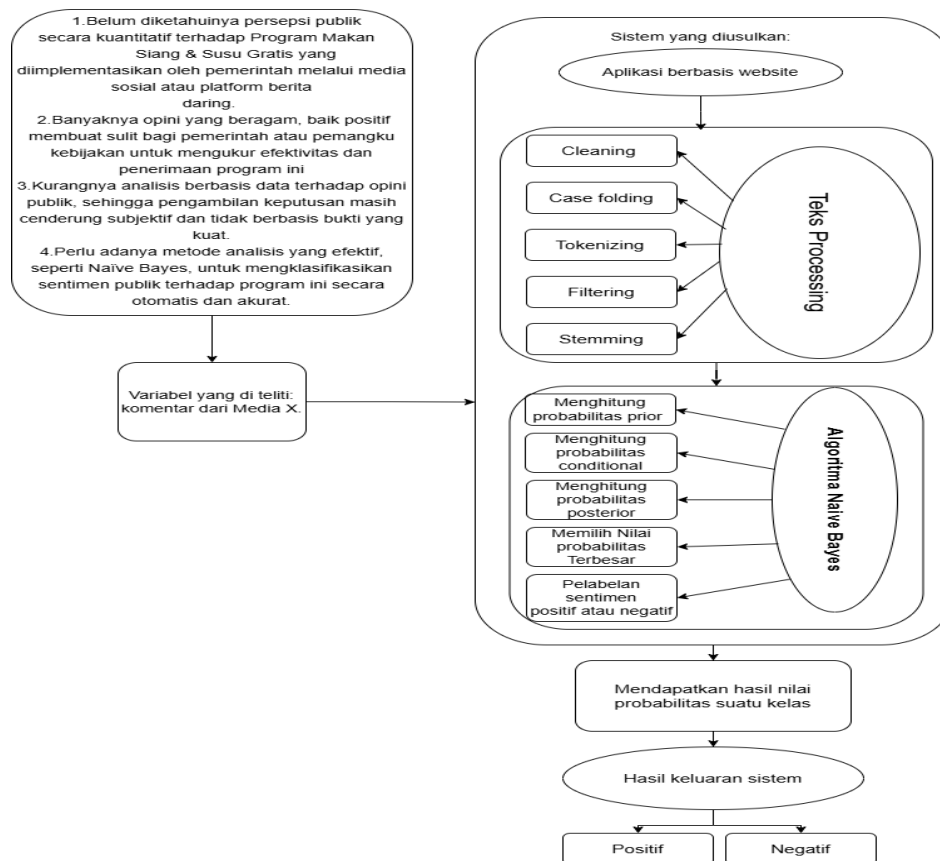
Siang dan Susu Gratis untuk lembaga pendidikan dan pesantren serta bantuan nutrisi untuk ibu hamil dan bayi, guna mengatasi stunting kondisi kekurangan nutrisi kronis yang berdampak fisik dan kognitif (Saputra & Hasan, 2024). Program ini menimbulkan kontroversi dan menjadi topik hangat di platform X, memicu diskusi publik dengan beragam opini. Dalam konteks ini, analisis sentimen digunakan untuk mempelajari opini dan sikap masyarakat terhadap isu tertentu (Nursingah et al., 2024). Data yang telah diproses akan dikonversi ke format numerik menggunakan metode TF-IDF, yang memberi bobot pada kata berdasarkan frekuensi kemunculannya dalam dokumen (Deolika & Taufiq Luthfi, 2021). Naïve Bayes Classifier digunakan dalam penelitian ini karena pendekatan probabilitiknya yang efektif dalam analisis teks. Hal ini membuatnya cocok untuk analisis sentimen di media sosial, di mana data yang diperoleh mencakup berbagai topik, masalah, dan perbedaan penggunaan bahasa (Mauliza & Sipayung, 2024). Kecepatan dan kesederhanaan algoritma Naïve Bayes ideal dipilih untuk menganalisis data dalam jumlah besar seperti yang dihasilkan oleh ulasan media sosial, dalam hal ini Twitter (Saputra & Hasan, 2024).

Studi ini menganalisis data dari rentang waktu 10 Februari 2024 sampai dengan 20 Maret 2025. Periode ini dipilih untuk mencakup dinamika opini publik menjelang dan setelah pemilihan presiden, serta selama pelaksanaan program makan siang gratis untuk memahami respons publik terhadap program tersebut di platform X. Data dikumpulkan melalui perintah tweet-harvest di lingkungan Google Colab, menggunakan token autentikasi dan parameter pencarian tertentu. Hasilnya disimpan dalam format CSV untuk keperluan analisis. Hasilnya diharapkan memberikan gambaran yang lebih jelas mengenai penerimaan masyarakat terhadap program makan siang gratis.

II. METODOLOGI

Penelitian ini menggunakan pendekatan machine learning dengan menerapkan algoritma Naïve Bayes untuk mengklasifikasikan sentimen dari teks media sosial, khususnya platform X (Twitter). Gambar 1 menunjukkan alur pemikiran yang digunakan dalam penelitian, dimulai dari identifikasi masalah,

pengumpulan data, pemrosesan teks, klasifikasi, hingga evaluasi model. Setiap tahapan dijelaskan secara rinci pada bagian berikut.



Gambar 1. Alur Penelitian

Machine Learning adalah proses di mana komputer belajar dari data tanpa instruksi eksplisit, memungkinkan sistem mengenali pola dan melakukan generalisasi untuk menyelesaikan masalah secara mandiri (Pipin et al., 2024). Penelitian ini menggunakan machine learning untuk menganalisis opini publik terkait program makan siang gratis dari data media sosial. Text mining adalah proses berbasis pengetahuan yang melibatkan interaksi pengguna dengan kumpulan dokumen dalam periode tertentu menggunakan berbagai metode analisis (Findawati et al., 2020). Teknik ini cocok digunakan dalam penelitian berbasis data teks seperti ulasan atau opini publik di media sosial.

Data dikumpulkan secara otomatis dari Twitter menggunakan pemrograman di Google Colab melalui pustaka tweet-harvest dan autentikasi token. Setelah data dikumpulkan, tahap berikutnya adalah pra-pemrosesan teks yang bertujuan untuk

mempersiapkan teks menjadi data yang akan diproses lebih lanjut (Supriyanto et al., 2023). Cleaning adalah proses menghapus karakter atau elemen yang tidak relevan dari teks seperti tanda baca, simbol, atau link (Asriana et al., 2024). Proses ini mengubah semua huruf menjadi huruf kecil untuk menjaga konsistensi teks (Indra Buana & Brahma Arianto, 2024). Tokenisasi adalah proses memecah kalimat menjadi unit-unit kecil berupa token kata (Indra Buana & Brahma Arianto, 2024). Normalisasi dilakukan untuk mengubah kata tidak baku menjadi bentuk baku atau standar (Asriana et al., 2024). Filtering adalah proses menghapus kata-kata umum seperti "yang", "dan", "atau" yang tidak memiliki kontribusi besar terhadap analisis (Indra Buana & Brahma Arianto, 2024). Stemming bertujuan mengembalikan kata ke bentuk dasarnya, misalnya "bermain" menjadi "main" (Indra Buana & Brahma Arianto, 2024). Setelah proses pra-pemrosesan selesai, data teks dikonversi menjadi representasi numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). TF-IDF digunakan untuk mengukur tingkat kepentingan suatu kata terhadap dokumen tertentu dalam keseluruhan korpus data (Sera et al., 2023). Metode ini terdiri dari tiga tahapan utama: perhitungan *Term Frequency* (TF), *Document Frequency* (DF), dan *Inverse Document Frequency* (IDF) (Gerald Halim & Noer, 2023). Dengan TF-IDF, setiap kata dikonversi ke dalam bobot numerik, sehingga bisa digunakan dalam proses klasifikasi selanjutnya (Junaedi et al., 2024). Tahap ini dimulai dengan membagi data menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk membangun model klasifikasi, sementara data uji digunakan untuk mengukur performa model dalam mengelompokkan data baru (Rahel Lina Simanjuntak et al., 2023). Penelitian ini menggunakan algoritma Naïve Bayes karena pendekatan probabilitiknya yang efisien dalam menganalisis data teks, dengan menghitung peluang kelas berdasarkan frekuensi kata. (Christian & Fenriana, 2023). Metode *confusion matrix* digunakan untuk mengevaluasi akurasi model klasifikasi berdasarkan jumlah prediksi benar dan salah pada tiap kelas.

III. HASIL DAN PEMBAHASAN

Aplikasi analisis sentimen berbasis web menunjukkan performa yang sebanding dengan RapidMiner. Berdasarkan Tabel 1, kedua platform mencatatkan

akurasi 79,72%. Selisih presisi, recall, dan F1-score pun kecil, yaitu masing-masing 0,63%, 5,28%, dan 0,21%. Hal ini menunjukkan bahwa model klasifikasi pada aplikasi memiliki validitas tinggi dan mampu mereplikasi hasil dari alat bantu standar.

Tabel 1. Hasil Perhitungan Confusion Matrix

	Akurasi	Presisi	Recall	F-1score
RapidMiner	79,72%	90,95%	86,36%	88,59%
Aplikasi Website	79,72%	97,08%	80,98%	88,3%

Aplikasi terbukti memenuhi verifikasi fungsional dan matematis, memperkuat keandalan Naive Bayes dalam klasifikasi sentimen ulasan.

Analisis Sentimen

Dashboard

Pelabelan Leksikon

Klasifikasi Naive Bayes

Hasil Pelabelan Leksikon

Teks	Label
full_text	positif
Belum dilantik sudah ada potensi ingkar janji : 1.Makan Siang Gratis 2.Kelanjutan Program Jokowi 3.DLL https://t.co/KSLvZAI5fI	positif
@HumorJonTampam Program makan siang gratis	positif
@rwhlyudis cukup buat program makan siang gratis bang	positif
@TioWii @gebutahupedes @Heraloebss Siapa yg mental gratisan ? Sekarang siapa yg bertanggung jawab program MAKAN SIANG GRATIS tersebut jika terjadi sesuatu Jgn lupa tragedi ini Banyak yg menolok fucksin Karena talu konsekuensi fucksin Akhirnya tidak ada pe	negatif
@Nasikebully11 @Heraloebss Sekarang siapa yg mengelola program MAKAN SIANG GRATIS tersebut ?? Pejabat siapa ?? Harus tahu donk publik Tokoh pejabat siapa yg mengelola program tersebut Karena baginda raja @jokowi dynasty sudah memberikan restu untuk progra	positif
@Nasikebully11 @Heraloebss Siapa yang mengelola program MAKAN SIANG GRATIS tersebut?? Pejabat siapa ? Dan struktur ya siapa ? Sponsor dana ya siapa ? Jika terjadi sesuatu pada program tersebut Yang akan bertanggung jawab Jangan jadi ambigu apalagi lepas ta	negatif
@ipad10gen @Heraloebss Sekarang saya tanya ke anda Pejabat siapa yang akan mengelola program MAKAN SIANG GRATIS tersebut?? Agar publik tahu dan minta pertanggung jawaban jika terjadi sesuatu pada program tersebut	positif
@ipad10gen @Heraloebss Sekarang siapa yg akan tanggung jawab jika terjadi sesuatu pada program MAKAN SIANG GRATIS tersebut ??? Jangan lupa tragedi PT CONGORNIA 19 Fucksin membuat anak SD Indonesia menderita Dan tidak ada satupun pejabat negeri ini bertanggung	negatif

Gambar 2. Tampilan Antarmuka Hasil Leksikon

Antarmuka ini menyajikan tabel dengan dua kolom: *Teks* (data teks mentah) dan *Label* (hasil pelabelan leksikon).

Analisis Sentimen

Dashboard

Pelabelan Leksikon

Klasifikasi Naive Bayes

Hasil Klasifikasi

Akurasi (Split 80:20): 79.72%

Confusion Matrix

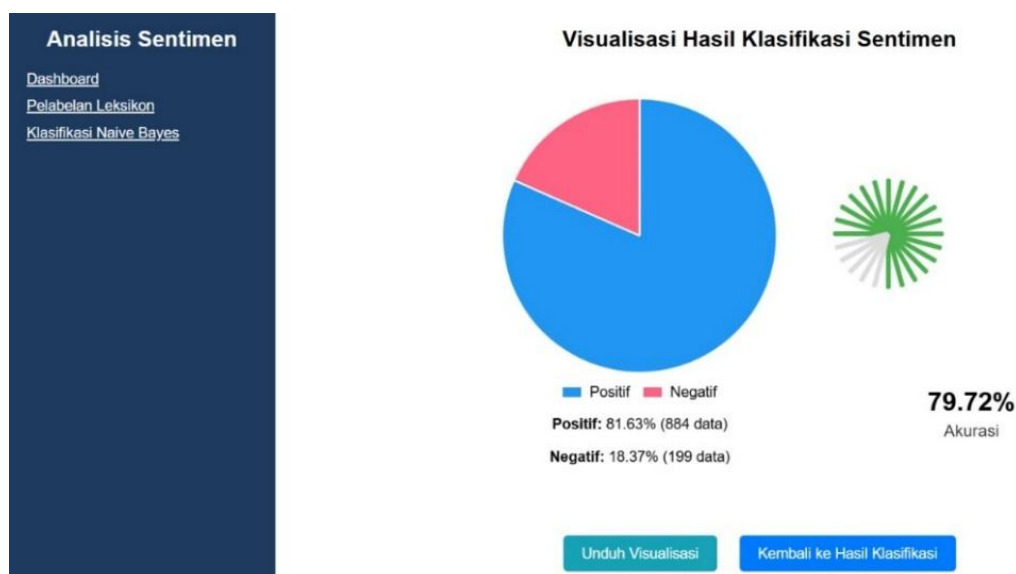
	Prediksi Positif	Prediksi Negatif
Aktual Positif	166	39
Aktual Negatif	5	7

Tabel Hasil Prediksi (Data Test)

Teks Asli	Hasil Preprocessing	Label Aktual	Label Prediksi
@kompascom Koq Bukan Untuk Program Makan Siang Gratis Dan IKN. Kalau Iya napa?;;	program makan siang gratis	positif	negatif
Program makan siang gratis Prabowo-Gibran disebut membutuhkan beras 6 7 juta ton beras per tahun. Ini belum angka fixed! https://t.co/1iGNSR2Csb. Sementara itu produksi beras nasional saat ini fluktuatif. Pada 2023 diperkirakan sebesar 30 90 juta ton. https://t.co/SgFrhq0Su5 https://t.co/s2DzL0QFFK;;	program makan siang gratis prabowogibran butuh beras beras angka fixed produksi beras nasional fluktuatif	positif	positif
@vwexcellent Kalo ngeliat program makan siang gratis yang diadakan di sana. Udah bener milih 02. Kebijakan buat mereka sekarang lebih bermanfaat daripada sibuk ngurusin shitposting gak jelas yang udah gak relevan	ngeliat program makan siang gratis bener milih bijak manfaat sibuk ngurusin shitposting	positif	positif

Gambar 3. Tampilan antarmuka Hasil klasifikasi

Hasil analisis sentimen ditampilkan berdasarkan klasifikasi algoritma Naive Bayes terhadap dataset yang diunggah.



Gambar 4. Tampilan Antarmuka Visualisasi

Visualisasi hasil berupa grafik lingkaran yang menggambarkan proporsi sentimen positif dan negatif berdasarkan hasil klasifikasi.

IV. SIMPULAN

Penelitian ini menghasilkan aplikasi web untuk analisis sentimen publik terhadap Program Makan Siang & Susu Gratis, menggunakan algoritma Naive Bayes dan representasi data berbasis frekuensi term (TF). Algoritma Naive Bayes terbukti efektif dalam mengklasifikasikan teks berbahasa Indonesia setelah melalui tahapan prapemrosesan seperti case folding, tokenisasi, pembersihan karakter, penghapusan stopwords, dan transformasi frekuensi. Dengan skema data 80% pelatihan dan 20% pengujian, evaluasi menggunakan confusion matrix menunjukkan akurasi 79,72%, presisi 97,08%, recall 80,98%, dan F1-score 88,3%, yang mencerminkan kinerja sistem yang baik dalam mengidentifikasi sentimen positif dan negatif.

DAFTAR PUSTAKA

Asriana, S., Oktariani Nurul Pratiwi, & Hanif Fakhurroja. (2024). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Support Vector Machine

(SVM). *Analisis Sentimen Pada Media Sosial Twitter Menggunakan Support Vector Machine (SVM)*, 11.

Christian, R., & Fenriana, I. (2023). *PERANCANGAN APLIKASI ANALISIS SENTIMEN PADA ULASAN MASKAPAI PENERBANGAN DI TRIPADVISOR MENGGUNAKAN METODE NAÏVE BAYES*.

Deolika, A., & Taufiq Luthfi, E. (2021). ANALISIS PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING. *Jurnal Teknologi Informasi*, 3(2).

Findawati, O. Y., Muhammad, M. M., Rosid, A., Kom, S., & Kom, M. (2020). *BUKU AJAR TEXT MINING Diterbitkan oleh UMSIDA PRESS*.

Gerald Halim, & Noer, R. (2023). *IMPLEMENTASI ALGORITMA NAÏVE BAYES DAN TF-IDF DALAM ANALISIS SENTIMEN DATA ULASAN (Studi Kasus: Ulasan Review Aplikasi E-Commerce Shopee Di Situs Google Playstore)*.

Indra Buana, M., & Brahma Arianto, D. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi ZenPro dengan Implementasi Algoritma Support Vector Machine (SVM). *Adopsi Teknologi Dan Sistem Informasi (ATASI)*, 3(1), 45–52. <https://doi.org/10.30872/atasi.v3i1.1092>

Junaedi, Hendra Gunawan, A., Kuswanto, V., & Jonathan. (2024). Tinjauan Support Vector Machine dalam Text-Mining untuk Analisis Sentimen di Sektor Pariwisata. *Bit-Tech*, 7(2), 323–330. <https://doi.org/10.32877/bt.v7i2.1810>

Mauliza, R. N., & Sipayung, Y. R. (2024). Penerapan Text Mining Dalam Menganalisis Pendapat Masyarakat Terhadap Pemilu 2024 Pada Media Sosial X Menggunakan Metode Naive Bayes. *Technomedia Journal*, 9(1), 1–16. <https://doi.org/10.33050/tmj.v9i1.2212>

Nursingah, 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>

- Pipin, S. J., Judijanto, L., & Reba, F. (2024). *ALGORITMA PEMBELAJARAN MESIN (Dasar, Teknik, dan Aplikasi)*.
<https://www.researchgate.net/publication/379479664>
- Rahel Lina Simanjuntak, Theresia Romauli Siagian, Vina Anggriani, & Arnita Arnita. (2023). Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 3(3), 23–39. <https://doi.org/10.55606/teknik.v3i3.2411>
- Saleh, A., & Furbani, W. (2022). LITERASI KOMUNIKASI WARGANET GENERASI MILENIAL DI MEDIA SOSIAL (STUDI KASUS WARGANET DI KOTA MATARAM). In *Jurnal Ilmu Perpustakaan (JIPER) FISIP UMMAT* | (Vol. 4, Issue 2). <https://tekno.kompas.com/read/2021/02>
- 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>
- Sera, E., Komputer, S., Makassar, H., & Riset dan Inovasi, B. (2023). *Analisis Sentimen Ulasan Produk di E-Commerce Bukalapak Menggunakan Natural Language Processing Sentiment Analysis of Product Reviews on E-Commerce's Bukalapak Using Natural Language Processing*. <https://www.kaggle.com/>.
- Supriyanto, J., Alita, D., & Isnain, A. R. (2023). Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 4(1), 74–80. <https://doi.org/10.33365/jatika.v4i1.2468>