

Analisis Sentimen Opini Publik Program Makan Siang Gratis dengan *Random Forest* Pada Media X

Muhamad Azhari*, Parjito

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ^{1,*}muhamad_azhari@teknokrat.ac.id, ²djito@teknokrat.ac.id

Email Penulis Korespondensi: muhamad_azhari@teknokrat.ac.id

Submitted: 09/12/2024; Accepted: 25/12/2024; Published: 26/12/2024

Abstrak– Program "Makan Siang Gratis" yang menjadi bagian dari kampanye Pemilu Indonesia 2024 telah menjadi perbincangan hangat di media sosial, khususnya di *platform X*. Program ini bertujuan untuk meningkatkan status kesehatan dan gizi anak-anak serta menekan angka stunting melalui penyediaan makan siang gratis dan susu bagi anak-anak serta ibu hamil. Penelitian ini menganalisis sentimen masyarakat terhadap program tersebut menggunakan algoritma *Random Forest*. Data yang dianalisis berasal dari 9.347 tweet yang diperoleh melalui teknik *crawling*. Sebagian besar sentimen yang ditemukan bersifat negatif (8.021 data), sedangkan sentimen positif hanya berjumlah 430 data. Tahapan *preprocessing* meliputi pembersihan data, *case folding*, tokenisasi, penghapusan *stopword*, dan *stemming*. Ketidakseimbangan data antara sentimen positif dan negatif diatasi dengan teknik *Synthetic Minority Over-sampling Technique (SMOTE)*, yang menghasilkan distribusi data yang lebih merata. Sebelum penerapan *SMOTE* model analisis mencapai akurasi 96%. Dan setelah penerapan *SMOTE*, model analisis mencapai akurasi 100%, dengan peningkatan signifikan pada *precision*, *recall*, dan *F1-Score*. Hasil analisis menunjukkan bahwa sentimen positif terutama berkaitan dengan manfaat program ini dalam aspek kesehatan dan pendidikan, sedangkan sentimen negatif menyoroti kritik terhadap implementasi dan penggunaan anggaran. Penelitian ini menunjukkan pentingnya analisis sentimen dalam mengevaluasi program sosial dan memahami opini publik.

Kata Kunci: Program Makan Siang Gratis, Media Sosial, Sentimen Masyarakat, *Random Forest*, *SMOTE*.

Abstract–The "Free Lunch Program," introduced as part of the 2024 Indonesian election campaign, became a hot topic on social media, especially on the platform X. This program aims to improve children's health and nutrition while reducing stunting rates by providing free lunches and milk to children and pregnant women. A study was conducted to analyze public sentiment regarding the program using the *Random Forest* algorithm. The data consisted of 9,347 tweets collected through web crawling. The analysis revealed that the majority of sentiments were negative (8,021 entries), while positive sentiments accounted for only 430 entries. The preprocessing steps included data cleaning, case folding, tokenization, stopword removal, and stemming. The imbalance between positive and negative sentiment data was addressed using the *Synthetic Minority Over-sampling Technique (SMOTE)*, resulting in a more balanced dataset. After applying *SMOTE*, the model achieved 100% accuracy, with significant improvements in *precision*, *recall*, and *F1-Score*. The analysis showed that positive sentiments focused on the program's health and educational benefits, while negative sentiments highlighted criticism of implementation and budget allocation. This study demonstrates the value of sentiment analysis in evaluating social programs and understanding public perceptions.

Keywords: Free Lunch Program, Social Media, Public Sentiment, *Random Forest*, *SMOTE*.

1. PENDAHULUAN

Masyarakat saat ini menghadapi kemajuan pesat dalam bidang teknologi informasi dan komunikasi, yang mendorong penggunaan media sosial secara luas. Pada Februari 2024, Indonesia akan menyelenggarakan pemilihan umum yang diadakan setiap lima tahun. Dari tiga pasangan calon yang bersaing, pasangan dengan nomor urut 02 menawarkan program unggulan berupa penyediaan makan siang gratis sebagai salah satu daya tarik utama kampanyenya [1]. Program makan siang gratis adalah salah satu upaya yang dirancang untuk meningkatkan kesehatan serta status gizi anak-anak di berbagai jenjang pendidikan, mulai dari prasekolah hingga sekolah menengah atas (SMA) dan pesantren. Selain menyediakan makan siang gratis, program ini juga memberikan susu kepada anak-anak dan ibu hamil. Fokus utama dari inisiatif ini adalah memastikan terpenuhinya kebutuhan gizi, sekaligus menurunkan risiko stunting pada anak-anak sebuah isu kesehatan yang menjadi perhatian serius di Indonesia [2].

Dengan menawarkan makan siang gratis, pasangan Prabowo–Gibran berupaya memberikan perlindungan sosial bagi kelompok rentan yang menghadapi risiko kemiskinan dan ketidakadilan. Inisiatif ini juga bertujuan memastikan bahwa seluruh lapisan masyarakat dapat berkontribusi dalam pembangunan nasional dan menikmati manfaat dari kemajuan bangsa [3]. Melalui janji tersebut, Prabowo dan Gibran menegaskan dedikasi mereka untuk meningkatkan kesejahteraan masyarakat serta mengurangi kesenjangan ekonomi yang ada di Indonesia [4].

Pembahasan mengenai program tersebut mendapat beragam tanggapan dari netizen (pengguna), termasuk salah satu respons dari pengguna platform media sosial X [5]. Media X adalah *platform* media sosial yang sangat populer di Indonesia. X dapat dimanfaatkan untuk menyampaikan opini dan pandangan seseorang [6]. Saat ini, perkembangan teknologi berjalan dengan sangat cepat. Seiring berjalannya waktu media sosial berperan penting sebagai *platform* bagi pengguna untuk menyampaikan pendapat, yang dapat berupa pujian, kritik, ujaran kebencian, atau bahkan hoax, yang semuanya memiliki potensi untuk memicu perdebatan. Salah satu *platform* tersebut adalah media X, di mana pengguna memiliki kebebasan untuk mengekspresikan pandangan mereka [7]. Banyaknya opini yang beredar di media sosial terkait program ini menunjukkan perlunya melakukan analisis mendalam terhadap respons publik [1].

Dalam konteks ini, analisis sentimen masyarakat terhadap program makan siang gratis menjadi sangat penting untuk memahami persepsi dan respons mereka terhadap inisiatif tersebut. Analisis sentimen berfokus pada data opini

subjektif yang tidak memiliki nilai objektif, melainkan bergantung sepenuhnya pada pandangan pribadi. Data tersebut bersumber dari individu-individu yang memiliki cara unik dalam menyampaikan pandangan serta preferensi masing-masing, sehingga menghasilkan keragaman perspektif [8].

Analisis sentimen adalah proses yang bertujuan untuk memahami, mengevaluasi, dan menentukan apakah suatu teks mengandung sentimen positif atau negatif. Proses ini memanfaatkan teknik analisis teks dengan algoritma komputasi yang bertugas mendeteksi pola serta makna dalam kalimat. Dengan demikian, analisis sentimen mampu mengidentifikasi emosi tersembunyi dalam opini, ulasan, atau komentar pengguna. Teknik ini memberikan wawasan berharga mengenai persepsi publik terhadap produk, layanan, atau kebijakan tertentu, sekaligus membantu dalam pengambilan keputusan berbasis data yang berasal dari interaksi pengguna di media sosial atau platform digital lainnya. Selain itu, analisis sentimen juga digunakan untuk memantau tren sosial dan opini publik, sehingga perusahaan atau institusi dapat meningkatkan mutu layanan dan produk mereka berdasarkan masukan dari Masyarakat [9].

Salah satu penelitian terkait menganalisis sentimen pengguna Twitter terhadap penggunaan pembayaran cashless dengan Shopeepay. Sentimen diklasifikasikan menjadi positif dan negatif menggunakan algoritma Random Forest. Dengan kedalaman tree 55 dan jumlah tree 300, penelitian ini mencapai precision 95%, recall 94%, F1-Score 95%, dan akurasi 95% [10]. Adapun penelitian lain menggunakan algoritma *Random Forest* dengan kedalaman tree 65 dan jumlah tree 400, menghasilkan performa terbaik dengan nilai *precision*, *recall*, dan *F1-Score* masing-masing 71%, serta akurasi 72%, menggunakan rasio data latih 90% dan data uji 10% [11]. Penelitian sebelumnya juga menggunakan algoritma *Random Forest* untuk analisis sentimen ulasan aplikasi Dana dengan membagi sentimen ke dalam tiga kategori: positif, negatif, dan netral. Penelitian ini mengevaluasi kinerja model berdasarkan indikator akurasi, presisi, *recall*, dan *F1-Score*. Pengujian dilakukan dengan memvariasikan jumlah *tree* dan kedalaman *tree* pada 1354 data, di mana setiap kelas memiliki 250 data. Dengan pembagian data latih dan uji sebesar 80%:20%, hasil menunjukkan nilai *precision*, *recall*, dan *F1-Score* masing-masing sebesar 84%, serta akurasi 84%, menggunakan kedalaman *tree* 65 dan jumlah *tree* 400 [12].

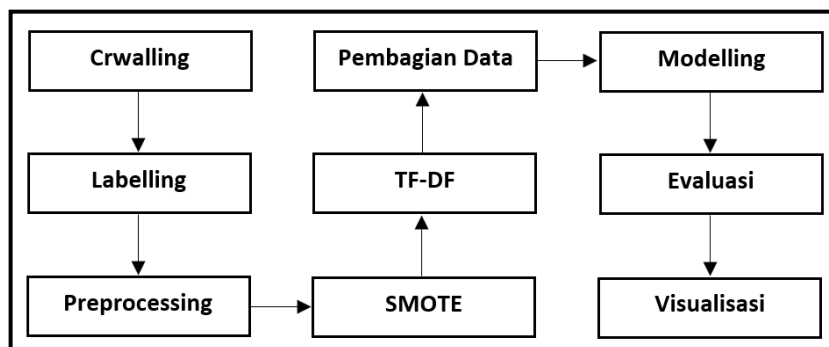
Penelitian serupa menggunakan algoritma Random Forest untuk menganalisis sentimen masyarakat terhadap vaksinasi COVID-19 di Twitter. Hasil penelitian menunjukkan bahwa sentimen masyarakat terhadap Vaksinasi Sinovac cenderung positif. Model ini mampu memprediksi sentimen tweet dengan akurasi 79%, *precision* 85%, *recall* 90%, dan *F1-score* 88% [13]. Berbagai penelitian telah menggunakan metode Random Forest untuk analisis sentimen, salah satunya meneliti sentimen komentar di YouTube terkait Islamofobia. Hasil analisis menunjukkan 631 komentar positif dan 369 komentar negatif atau mengandung unsur Islamofobia. Dalam eksperimen tersebut, algoritma *Random Forest* mencapai akurasi sebesar 79%, menunjukkan kinerjanya yang efisien dan cepat dalam mengklasifikasikan sentiment [14].

Berdasarkan permasalahan yang telah dijelaskan serta penelitian-penelitian yang relevan, penelitian ini bertujuan untuk mengklasifikasikan sentimen positif dan negatif dari komentar-komentar yang diposting di media sosial X menggunakan algoritma *Random Forest*. Dengan penelitian ini, diharapkan dapat diperoleh pemahaman yang lebih mendalam mengenai pandangan publik terhadap program makan siang gratis, serta untuk mengidentifikasi pola sentimen yang paling dominan. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pemahaman yang lebih sistematis mengenai cara opini masyarakat terhadap kebijakan pemerintah dapat diukur melalui analisis data media sosial, yang kemudian dapat digunakan untuk evaluasi dan perbaikan kebijakan di masa depan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Analisis Sentimen

Pada metode dalam penelitian ini dimulai dengan pengumpulan data terkait program makan siang gratis yang di ambil dari platform media X untuk mendapatkan opini publik. Seperti yang dijelaskan pada gambar 1, tahapan yang meliputi *crwaling* (pengumpulan data), *labelling*, *preprocessing* (*cleasing*, *case folding*, *tokenizing*, *stopwoard* dan *stemming*), SMOTE, TF-IDF, pembagian data, *modelling*, *evaluasi* dan terakhir visualisasi, apat dilihat pada gambar 1.



Gambar 1. Metode Penelitian

2.2 Pengumpulan Data

Penelitian ini mengumpulkan data menggunakan metode *crawling* dengan bantuan library *Harvest*. Metode ini memungkinkan pengambilan data dari media X dengan cara mengakses dan mengekstrak teks dari postingan maupun komentar [15]. Proses pengumpulan data dilakukan dengan menulis kode program yang menggunakan kata kunci untuk menemukan tweet berisi opini publik terkait program makan siang gratis sesuai preferensi yang ditentukan [16].

2.3 Pelebelan Data

Pelebelan data adalah proses memberikan tanda atau kategori pada data mentah dengan label tertentu untuk mendukung analisis atau pengembangan model. Proses ini menghasilkan skor sentimen yang digunakan untuk mengklasifikasikan komentar ke dalam kategori sentimen positif atau negatif. Tujuannya adalah untuk menilai dan menentukan orientasi sentimen yang terkandung dalam komentar yang sedang dianalisis [17]. Dapat dilihat pada persamaan satu

$$S_{Sentiment} = \begin{cases} \text{positive if } S_{positive} > S_{negative} \\ \text{negative if } S_{positive} < S_{negative} \end{cases} \quad (1)$$

Dalam konteks ini, $S_{sentiment}$ mengacu pada kalimat yang berisi sentimen dalam dataset, baik itu positif maupun negatif. Sementara itu, *positive* adalah kalimat dalam dataset yang mengandung sentimen positif, sedangkan *negative* adalah kalimat yang mencerminkan sentimen negatif. Ketiga istilah ini berguna dalam proses klasifikasi dan analisis sentimen untuk mengidentifikasi jenis sentimen yang terkandung dalam data yang dikumpulkan.

2.4 Preprocessing

Proses penyaringan dan perbaikan data dikenal sebagai *preprocessing*. Pada tahap ini, data yang dikumpulkan sering kali berbentuk tidak terstruktur dan mengandung banyak elemen yang tidak relevan atau tidak sesuai [18]. Tahap *preprocessing* melibatkan pengolahan data hasil *crawling* sebanyak 8.451 data untuk persiapan analisis lebih lanjut. Proses ini mencakup beberapa teknik, antara lain *cleansing* (membersihkan data dari elemen yang tidak relevan), *case folding* (mengubah semua huruf menjadi bentuk kecil), *tokenizing* (memecah teks menjadi unit-unit kata), *stopword removal* (menghapus kata-kata umum yang tidak memiliki makna signifikan), *filtering* (menyaring data sesuai dengan kriteria tertentu), dan *stemming* (mengubah kata ke bentuk dasarnya) [19]. Dalam penelitian ini ada beberapa tahapan pada *preprocessing* sebagai berikut:

2.4.1 Cleansing

Data cleansing adalah proses penting dalam pengolahan data yang bertujuan untuk menghapus elemen-elemen yang tidak relevan atau tidak sesuai, seperti emoji, *hashtag*, karakter tertentu, angka, dan spasi. Langkah ini memastikan bahwa data yang akan digunakan menjadi lebih bersih, akurat, dan dapat diandalkan untuk analisis berikutnya. Dengan menghilangkan inkonsistensi serta informasi yang tidak diperlukan, analisis dapat menghasilkan hasil yang lebih valid dan efektif [20].

2.4.2 Case Folding

Case folding adalah langkah dalam pra-pemrosesan yang mengubah semua karakter dalam sebuah string ke format huruf tertentu, seperti seluruhnya menjadi huruf kecil atau huruf besar [21]. Proses ini hanya memproses huruf dalam rentang "a" hingga "z," sehingga karakter di luar rentang tersebut akan dihapus [22].

2.4.3 Tokenizing

Tokenizing, atau proses tokenisasi, adalah langkah memecah suatu kalimat menjadi unit-unit kata individu. Proses ini melibatkan pemisahan seluruh karakter dalam sebuah dokumen menjadi komponen-komponen kecil yang disebut token, yang mewakili kata-kata atau bagian-bagian lain dalam teks [19].

2.4.4 Stopword

Stopword adalah metode untuk menghilangkan kata-kata yang dianggap terlalu umum atau kurang relevan dalam suatu teks atau dokumen [23]. *Stopword* merujuk pada kata-kata umum yang sering muncul dalam berbagai kalimat dalam jumlah banyak, seperti kata penghubung. Contohnya adalah kata-kata seperti "ke", "di", "dari", "adalah", dan "dan" [24].

2.4.5 Stemming

Stemming adalah langkah untuk mengubah kata-kata hasil penyaringan menjadi bentuk dasarnya dengan menghilangkan imbuhan yang melekat. Sebagai contoh, kata seperti "mempertahankan" diproses menjadi "tahan", "mengkui" berubah menjadi "akui", dan "melindungi" menjadi "lindung" [25].

2.5 Pembagian Data

Langkah selanjutnya dalam adalah pembagian data dimana memisahkan data menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Penelitian ini menggunakan empat proporsi pembagian data, yaitu 50:50, 90:10, 70:30, dan 80:20. Setelah pembagian dilakukan, data pelatihan digunakan untuk membangun model, sedangkan data pengujian dimanfaatkan untuk mengevaluasi performa model yang telah dilatih[26].

2.6 TF-IDF

TF-IDF adalah salah satu langkah penting yang digunakan sebelum melakukan klasifikasi untuk menghasilkan fitur melalui proses pembobotan kata. Metode ini bertujuan untuk meningkatkan akurasi dalam analisis data dengan memberikan nilai bobot tertentu pada setiap kata. Proses TF-IDF mencakup dua komponen utama: *Term Frequency* (TF), yang mengukur seberapa sering sebuah kata muncul dalam dokumen, dan *Inverse Document Frequency* (IDF), yang menilai seberapa jarang kata tersebut muncul di seluruh kumpulan dokumen. Gabungan kedua nilai ini membantu menentukan tingkat relevansi sebuah kata terhadap dokumen tertentu [27]. Pembobotan ini kemudian diterapkan sebagai strategi efektif dalam klasifikasi dokumen, seperti yang dijelaskan dalam Persamaan 2.

$$TF * IDF (d,t) = TF (d,t) * \log \frac{N}{df(t)} \quad (2)$$

Dalam konteks ini, TF atau *Term Frequency* adalah nilai yang menunjukkan seberapa sering suatu term *t* muncul dalam dokumen tertentu. IDF(d,t) atau *Inverse Document Frequency* mengukur seberapa penting term tersebut dalam seluruh kumpulan dokumen. Nilai ini dihitung dengan mempertimbangkan jumlah total dokumen N dan jumlah dokumen yang mengandung term tertentu, yang dilambangkan dengan Df(t). Sementara itu, (d,t) merepresentasikan jumlah kemunculan term t dalam dokumen tertentu d. Gabungan dari TF dan IDF digunakan untuk menilai relevansi term dalam analisis teks atau informasi yang diolah dalam model pencarian dan pengolahan bahasa alami.

2.7 SMOTE

SMOTE (*Synthetic Minority Over-sampling Technique*) merupakan metode efektif yang sering digunakan untuk menangani masalah ketidakseimbangan data. Dalam konteks analisis sentimen, perbedaan jumlah data antara sentimen positif dan negatif dapat menyebabkan model lebih fokus pada kelas yang lebih dominan. Untuk mengatasi hal ini, SMOTE bekerja dengan menghasilkan sampel sintetis pada kelas minoritas, sehingga distribusi data menjadi lebih merata dan model dapat memberikan performa yang lebih adil dan akurat [28].

2.8 Pembangun Mode Random Forest

Random Forest adalah teknik yang digunakan untuk klasifikasi dan regresi dengan memanfaatkan metode pembelajaran ansambel. Dalam pendekatan ini, sejumlah pohon keputusan (*decision tree*) digunakan sebagai pengklasifikasi dasar (*base classifier*), yang kemudian dikembangkan dan digabungkan untuk menghasilkan prediksi akhir. Metode ini memanfaatkan *bootstrap sampling* untuk membuat setiap pohon, di mana sejumlah prediktor dipilih secara acak saat melakukan prediksi. Prediksi akhir diperoleh dengan menggabungkan hasil dari semua pohon: untuk klasifikasi, proses ini dilakukan melalui suara mayoritas (*majority voting*), sedangkan untuk regresi menggunakan perhitungan rata-rata hasil prediksi[29]. Dapat menggunakan persamaan 3.

$$fx = Average(f_1(x), f_2(x), \dots, f_n(x)) \quad (3)$$

Dalam hal ini, f(x) adalah hasil prediksi akhir dari model berdasarkan input x. Sementara itu, f1-n adalah hasil prediksi dari masing-masing pohon keputusan ke-n dalam ensemble seperti *Random Forest*. Hasil akhir f(x) diperoleh dengan menggabungkan semua prediksi dari f1-n melalui proses voting atau kombinasi, yang bertujuan meningkatkan akurasi model dibandingkan dengan prediksi pohon individu.

2.9 Evaluasi Model

Tahapan evaluasi dalam analisis sentimen terdiri dari beberapa langkah utama. Pertama, dataset yang telah diproses dibagi menjadi data latih dan data uji. Data pelatihan digunakan untuk melatih model sentimen, sementara data uji digunakan untuk mengukur kinerjanya. Selanjutnya, metrik evaluasi seperti akurasi, presisi, recall, dan *F1-score* digunakan untuk menilai kemampuan model dalam mengklasifikasikan sentimen secara akurat. Hasil evaluasi ini dianalisis lebih lanjut untuk mengidentifikasi keunggulan, kelemahan, dan peluang peningkatan kinerja model[30]. Pada penelitian ini, evaluasi dilakukan menggunakan metode *confusion matrix*. Data uji dikelompokkan ke dalam empat kategori: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Akurasi dihitung sebagai rasio antara jumlah total nilai *True* dengan keseluruhan data[31]. Setiap elemen dalam *Confusion Matrix* yang direpresentasikan pada persamaan berikut.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4)$$

$$Presisi = \frac{TP}{TP+FP} \quad (5)$$



$$Recal = \frac{TP}{TP+FN} \tag{6}$$

$$f1 - score = \frac{2xPresisixRecall}{PresisixRecall} \tag{7}$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dilakukan melalui metode *crawling* dari media sosial Twitter. Kata kunci yang dipakai dalam pengumpulan data ini antara lain “Program Makan Siang Gratis” dan “Program Bermanfaat.” Sebanyak 9.347 *record* berhasil dikumpulkan, dengan data mencakup periode tahun 2024. Data tersebut kemudian disimpan dalam format Excel untuk keperluan pengolahan dan analisis sentimen lebih lanjut. Proses *crawling* dilakukan menggunakan Google Colab dengan bahasa pemrograman Python, memanfaatkan *library* *pandas* untuk membantu proses pengambilan data. Dataset yang digunakan untuk pemodelan dapat dilihat dalam Tabel 1.

Tabel 1. Sampel Hasil Pengumpulan Data

Username	Tweet
nabiylarisfa	Inggris juga punya program makan siang gratis untuk anak sekolah. Tapi terbagi jadi dua: 1) infant free lunch program buat anak reception year 1 dan year 2 semua dapet. 2) school free lunch buat year 3 - year 12 cuma dikasih untuk yg keluarganya kategori ekonomi bawah. Program Makan Siang Gratis Berdampak Positif pada Peningkatan Asupan Protein Anak @viaaan11 @sayurebung @anonjawa Anggarannya bisa dialokasikan ke program yang lebih bermanfaat. Stunting itu sejak dalam kandungan ngasih anak sd makan siang gratis itu itungannya udah telat. Lebih baik anggarannya dijor-jorin ke program pra kehamilan dan saat kehamilan si ibu. Manfaatnya jauh lebih luas
Nan4_ecia	

3.2 Labelling Data

Dalam penelitian ini, skor polaritas dihitung berdasarkan jumlah kata yang diidentifikasi menggunakan kamus leksikon, sehingga sentimen dapat diklasifikasikan ke dalam dua kategori, yaitu positif dan negatif. Tahap berikutnya adalah memverifikasi hasil sentimen berdasarkan skor polaritas yang diperoleh. Proses pelabelan sentimen ini dikelompokkan ke dalam dua kelas: sentimen positif dan sentimen negatif. Hasil analisis sentimen menunjukkan bahwa dari seluruh data yang dianalisis, terdapat 430 data yang termasuk dalam kategori positif, sementara 8.021 data masuk dalam kategori negatif. Rincian pelabelan data untuk program makan siang gratis ini disajikan pada Tabel 2.

Tabel 2. Hasil Labelling Data

Tweet	Label
inggris program makan siang gratis anak sekolah bagi infant free lunch program anak reception year year dapet school free lunch year year kasih yg kategori ekonomi	Positif
viaaan anonjawa program yang stunting kandung ngasih anak sd makan siang gratis udah telat program pra hamil hamil si luas	Negatif

3.3 Preprocessing

Praprocessing atau *data preprocessing* adalah langkah penting dalam proses analisis data yang melibatkan penyusunan ulang data mentah menjadi bentuk yang lebih terstruktur dan bersih agar siap untuk dianalisis lebih lanjut. Tahapan ini bertujuan untuk meningkatkan mutu data, membuang komponen yang tidak relevan, serta memastikan data dapat diolah secara efisien oleh algoritma analitik. Selain itu, *preprocessing* membantu mengonversi data ke format yang lebih sederhana dan efektif untuk tahapan pemrosesan berikutnya. Berikut ini adalah beberapa langkah umum dalam preprocessing data untuk keperluan analisis sentimen.

3.3.1 Cleansing

Dalam analisis sentimen, *cleansing* adalah proses penting untuk membersihkan teks dari elemen-elemen yang tidak relevan atau dapat menghambat analisis. Langkah ini mencakup penghapusan tanda baca, karakter khusus, angka, tautan URL, sebutan seperti (@nama_pengguna), serta hashtag (#) yang umum di media sosial. Selain itu, kata-kata yang sering muncul namun tidak signifikan untuk analisis, seperti "di", "ke", atau "dari", juga dibuang. Proses ini juga melibatkan penghapusan spasi berlebih, memperbaiki kesalahan ketik, dan menyeragamkan variasi penulisan kata. Dengan cara ini, data menjadi lebih terstruktur dan seragam, mempermudah algoritma dalam mengidentifikasi sentimen secara akurat. Hasil dari proses *cleansing* ini dapat dilihat pada Tabel 3.

Tabel 3. Hasil Proses *Cleansing*

<i>Cleansing</i>
<p>Inggris juga punya program makan siang gratis untuk anak sekolah Tapi terbagi jadi dua infant free lunch program buat anak reception year dan year semua dapet school free lunch buat year year cuma dikasih untuk yg kategori ekonomi bawah</p> <p>viaaan anonjawa bisa me program yang lebih Stunting itu sejak dalam kandungan ngasih anak sd makan siang gratis itu udah telat Lebih baik ke program pra kehamilan dan saat kehamilan si ibu jauh lebih luas</p>

3.3.2 Case Folding

Case folding adalah langkah dalam *preprocessing* teks yang berfungsi mengonversi seluruh huruf dalam teks menjadi huruf kecil (*lowercase*). Tujuan dari proses ini adalah untuk menyelaraskan format teks agar perbedaan antara huruf besar dan kecil tidak memengaruhi hasil analisis. Misalnya, dalam kalimat "Makan SIANG Gratis", setelah melalui *casefolding*, teks akan menjadi "makan siang gratis". Proses ini meningkatkan konsistensi dan akurasi dalam analisis data, seperti dalam klasifikasi teks atau analisis sentimen, dengan menghindari kesalahan akibat perbedaan kapitalisasi. Hasil dari *case folding* dapat dilihat dari Tabel 4.

Tabel 4. Hasil Proses *Case Folding*

<i>Case Folding</i>
<p>inggris juga punya program makan siang gratis untuk anak sekolah tapi terbagi jadi dua infant free lunch program buat anak reception year dan year semua dapet school free lunch buat year year cuma dikasih untuk yg kategori ekonomi bawah</p> <p>viaaan anonjawa bisa me program yang lebih stunting itu sejak dalam kandungan ngasih anak sd makan siang gratis itu udah telat lebih baik ke program pra kehamilan dan saat kehamilan si ibu jauh lebih luas</p>

3.3.3 Tokenizing

Dalam analisis sentimen, *tokenizing* adalah proses memecah teks menjadi bagian-bagian kecil yang disebut token, seperti kata, frasa, atau karakter. Tujuan utama dari proses ini adalah mempermudah pengolahan teks dengan membagi kalimat atau paragraf menjadi unit-unit yang lebih sederhana, sehingga masing-masing token dapat dianalisis secara terpisah. Melalui *tokenizing*, algoritma dapat memahami konteks dari setiap kata serta mendeteksi pola emosi atau sentimen dalam teks. Misalnya, kalimat "program makan siang gratis" akan dipecah menjadi token seperti ['program', 'makan', 'siang', 'gratis']. Proses ini penting karena setiap token menyimpan informasi yang berguna untuk mengidentifikasi apakah teks tersebut memiliki sentimen positif atau negatif. Hasil dari proses *tokenizing* dapat dilihat dalam Tabel 5.

Tabel 5. Hasil Proses *Tokenizing*

<i>Tokenizing</i>
<p>['inggris', 'juga', 'punya', 'program', 'makan', 'siang', 'gratis', 'untuk', 'anak', 'sekolah', 'tapi', 'terbagi', 'jadi', 'dua', 'infant', 'free', 'lunch', 'program', 'buat', 'anak', 'reception', 'year', 'dan', 'year', 'semua', 'dapet', 'school', 'free', 'lunch', 'buat', 'year', 'year', 'cuma', 'dikasih', 'untuk', 'yg', 'kategori', 'ekonomi', 'bawah']</p> <p>['viaaan', 'anonjawa', 'bisa', 'me', 'program', 'yang', 'lebih', 'stunting', 'itu', 'sejak', 'dalam', 'kandungan', 'ngasih', 'anak', 'sd', 'makan', 'siang', 'gratis', 'itu', 'udah', 'telat', 'lebih', 'baik', 'ke', 'program', 'pra', 'kehamilan', 'dan', 'saat', 'kehamilan', 'si', 'ibu', 'jauh', 'lebih', 'luas']</p>

3.3.4 Stopword

Stopword adalah istilah untuk kata-kata umum yang sering muncul dalam teks, seperti "dan," "di," "yang," atau "maaf," yang memiliki nilai informasi rendah. Dalam *preprocessing* data teks, stopwords sering dihapus untuk meningkatkan efisiensi dan memastikan analisis lebih fokus pada kata-kata penting. Misalnya, kata "maaf," meskipun berguna dalam konteks tertentu, umumnya dianggap tidak banyak berkontribusi pada pemahaman makna keseluruhan dan biasanya termasuk dalam daftar *stopword* yang dihilangkan. Dalam analisis sentimen, menghapus *stopword* membantu algoritma lebih fokus pada kata-kata yang lebih signifikan dalam menentukan apakah suatu teks bersentimen positif atau negatif. Proses ini meningkatkan akurasi dan efisiensi dengan mengurangi interferensi dari kata-kata yang tidak relevan terhadap sentimen. Hasil penghapusan *stopword*, termasuk kata seperti "maaf," dapat dilihat dalam tabel 6.

Tabel 6. Hasil Proses *Stopword*

<i>Stopword</i>
<p>['inggris', 'program', 'makan', 'siang', 'gratis', 'anak', 'sekolah', 'terbagi', 'infant', 'free', 'lunch', 'program', 'anak', 'reception', 'year', 'year', 'dapet', 'school', 'free', 'lunch', 'year', 'year', 'dikasih', 'yg', 'kategori', 'ekonomi']</p> <p>['viaaan', 'anonjawa', 'program', 'yang', 'stunting', 'kandungan', 'ngasih', 'anak', 'sd', 'makan', 'siang', 'gratis', 'udah', 'telat', 'program', 'pra', 'kehamilan', 'kehamilan', 'si', 'luas']</p>

3.3.5 Stemming

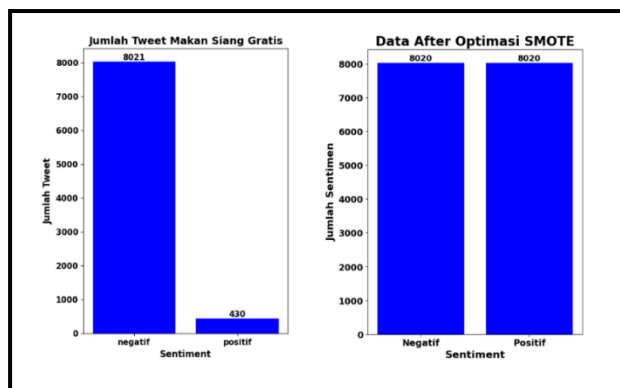
Stemming adalah teknik dalam *natural language processing* (NLP) yang bertujuan mengonversi kata menjadi bentuk dasar atau akarnya dengan menghapus imbuhan seperti awalan, sisipan, dan akhiran. Teknik ini digunakan untuk menyederhanakan teks sehingga kata-kata dengan makna serupa, meskipun memiliki variasi bentuk, dapat dikelompokkan bersama (contohnya: "berlari", "lari", dan "pelari" menjadi "lari"). *Stemming* membantu mengurangi kerumitan data teks, meningkatkan efisiensi algoritma analisis, dan memperbaiki hasil dalam aplikasi seperti klasifikasi teks, analisis sentimen, atau pencarian informasi dan dokumen yang dapat dilihat pada Tabel 7.

Tabel 7. Hasil Proses *Stemming*

<i>Stemming</i>
inggris program makan siang gratis anak sekolah bagi infant free lunch program anak reception year year dapat school free lunch year year kasih yg kategori ekonomi
viaaan anonjawa program yang stunting kandung ngasih anak sd makan siang gratis udah telat program pra hamil hamil si luas

3.4 Perbandingan Akurasi SMOTE

Dalam penelitian ini, hasil *preprocessing* pada analisis sentimen menunjukkan adanya 430 data dengan sentimen positif dan 8.021 data dengan sentimen negatif dari total 8.451 tweet. Ketimpangan antara jumlah data positif dan negatif ini menyebabkan algoritma *Random Forest* cenderung mempelajari pola dari data negatif. Akibatnya, model memiliki akurasi dan performa yang lebih tinggi dalam mengklasifikasikan sentimen negatif. Maka dari itu penilaian performa model tidak hanya didasarkan pada akurasi saja, tetapi juga menggunakan metrik lain dari *classification report*, seperti *precision*, *recall*, dan *F1-score* untuk menentukan efektivitas model secara keseluruhan. Untuk mengatasi ketidakseimbangan ini, penelitian menerapkan teknik optimasi *Synthetic Minority Over-sampling Technique* (SMOTE). Teknik ini bertujuan menyeimbangkan jumlah data positif (minoritas) agar setara dengan data negatif (mayoritas), sehingga model algoritma dapat mempelajari kedua jenis sentimen secara lebih seimbang.



Gambar 2. Hasil *Before After* SMOTE

Berdasarkan Gambar 2, terlihat perbandingan data sebelum dan sesudah penerapan optimasi menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE). Penerapan SMOTE meningkatkan jumlah data pada kategori sentimen positif dan negatif sehingga menjadi seimbang, dengan masing-masing kategori memiliki 8.020 data. Dengan adanya keseimbangan ini, model algoritma dapat mempelajari pola sentimen secara lebih efektif tanpa dipengaruhi ketimpangan antara data mayoritas dan minoritas.

Setelah proses penyeimbangan data, tahap selanjutnya adalah melatih dan menguji model algoritma *Random Forest*. Dalam penelitian ini, algoritma *Random Forest* menggunakan 70% dari total data untuk latih dan 30% untuk data uji. Data yang dianalisis mencakup kondisi sebelum dan sesudah penerapan SMOTE. Hasil dari model algoritma pada kedua kondisi tersebut kemudian dibandingkan untuk mengidentifikasi performa terbaik. Rincian lebih detail tentang hasil model dapat dilihat pada Gambar 3 dan 4.

```

==== Classification Report Random Forest Before SMOTE ====
precision    recall  f1-score   support

negatif      0.96    1.00    0.98    1597
positif      0.97    0.30    0.46     93

accuracy          0.96    1690
macro avg         0.96    0.65    0.72    1690
weighted avg     0.96    0.96    0.95    1690

==== Confusion Matrix Random Forest Before SMOTE ====
[[1596  1]
 [ 65  28]]
    
```

Gambar 3. Hasil Akurasi *Before* SMOTE

Dalam pembagian data dengan porsi 70% untuk pelatihan dan 30% untuk pengujian sebelum menerapkan optimasi SMOTE, algoritma *Random Forest* berhasil mencapai akurasi tinggi sebesar 96%. Namun, untuk menilai kemampuan model dalam membedakan sentimen positif dan negatif, akurasi saja belum memadai. Oleh sebab itu, performa model dievaluasi lebih mendalam menggunakan metrik seperti *precision*, *recall*, dan *F1-Score*.

```

==== Classification Report Random Forest After SMOTE ====
              precision    recall  f1-score   support

   negatif      1.00      1.00      1.00     1596
   positif      1.00      1.00      1.00     1612

 accuracy              1.00      3208
 macro avg              1.00      1.00      1.00     3208
 weighted avg          1.00      1.00      1.00     3208

 [[1595  1]
 [  0 1612]]
    
```

Gambar 4. Hasil Akurasi After SMOTE

Setelah menerapkan optimasi model menggunakan *Synthetic Minority Over-sampling Technique (SMOTE)*, terjadi peningkatan signifikan yang menghasilkan akurasi hingga 100%. Perbaikan ini terlihat jelas pada label negatif, dengan *precision* meningkat dari 96% menjadi 100%, *recall* tetap di 100%, dan *F1-Score* yang naik dari 98% ke 100%. Peningkatan serupa juga terjadi pada label positif, di mana *precision* naik dari 97% ke 100%, *recall* melonjak dari 30% menjadi 100%, dan *F1-Score* meningkat dari 46% menjadi 100%.

Tabel 8. Confusion Matrix Before SMOTE

Model	Prediction Class	Actual Class	
		Pred. Positif	Pred. Negatif
RF	Negatif	1596	1
	Positif	65	28

Dari hasil ini terlihat bahwa model memiliki performa yang sangat baik dalam memprediksi kelas negatif, namun performanya masih rendah untuk kelas positif. Hal ini menjadi alasan mengapa teknik SMOTE digunakan untuk menyeimbangkan data sehingga performa model dapat ditingkatkan untuk memprediksi kedua kelas secara lebih akurat.

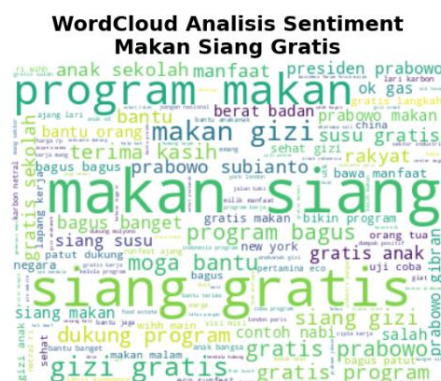
Tabel 9. Confusion Matrix After SMOTE

Model	Prediction Class	Actual Class	
		Pred. Positif	Pred. Negatif
RF	Negatif	1595	1
	Positif	0	1612

Setelah penerapan SMOTE, performa model meningkat secara signifikan dengan akurasi yang mencapai 100%. Prediksi untuk kedua kelas (negatif dan positif) menjadi jauh lebih seimbang, dengan kesalahan prediksi yang sangat minimal.

3.5 Visualiasi

Penelitian ini memanfaatkan *wordcloud* untuk menganalisis data teks secara efisien. Dalam *wordcloud*, kata-kata yang sering muncul ditampilkan dengan ukuran font yang bervariasi; semakin tinggi frekuensi kemunculan suatu kata, semakin besar ukurannya. Visualisasi *wordcloud* yang terkait dengan topik program makan siang gratis dapat dilihat sebagai berikut.



Gambar 5. Wordcloud Makan Siang Gratis

Grafik Gambar 8 di atas menunjukkan 10 kata dengan frekuensi tertinggi dalam tweet yang membahas "Makan Siang Gratis". Kata "makan" menduduki peringkat pertama dengan frekuensi tertinggi sebanyak 9.623 kali, diikuti oleh kata "gratis" sebanyak 7.936 kali, dan "program" dengan 7.658 kali. Kata-kata seperti "siang" (7.370 kali) dan "bantu" (5.234 kali) juga sering muncul, menggambarkan relevansi topik dengan inisiatif sosial. Kata "bagus" (2.654 kali) dan "Prabowo" (2.376 kali) mengindikasikan adanya apresiasi dan keterlibatan tokoh tertentu. Sementara itu, kata "gizi" (2.257 kali), "tidak" (2.065 kali), dan "anak" (1.434 kali) menunjukkan fokus pada aspek kesehatan dan target penerima program. Grafik ini mencerminkan perhatian publik terhadap inisiatif makan siang gratis, termasuk manfaat, pelaksana, serta tantangannya.

4. KESIMPULAN

Penelitian ini berhasil menganalisis sentimen masyarakat terhadap program "Makan Siang Gratis" dengan pendekatan pemrosesan teks berbasis *Natural Language Processing* (NLP). Melalui pengumpulan data sebanyak 9.347 tweet menggunakan metode *crawling* di Twitter, analisis ini menemukan bahwa sebagian besar sentimen yang diungkapkan bersifat negatif (8.021 data), sementara sentimen positif hanya mencapai 430 data. Proses *preprocessing*, meliputi *cleansing*, *case folding*, *tokenizing*, penghapusan *stopword*, dan *stemming*, memastikan data lebih bersih dan terstruktur untuk analisis lebih lanjut. Ketidakseimbangan data antara sentimen positif dan negatif diatasi menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), yang meningkatkan distribusi data menjadi seimbang. Teknik ini memungkinkan model Random Forest untuk mempelajari pola sentimen positif dan negatif secara lebih efektif. Sebelum penerapan SMOTE, model memiliki akurasi sebesar 96% dengan performa rendah pada prediksi sentimen positif. Setelah penerapan SMOTE, akurasi meningkat menjadi 100% dengan peningkatan signifikan pada semua metrik evaluasi, seperti *precision*, *recall*, dan *F1-Score*, yang mencapai 100% pada kedua kelas sentimen. Visualisasi *wordcloud* mengungkap bahwa kata-kata seperti "makan siang," "gratis," dan "program" menjadi fokus utama diskusi. Sentimen positif mencerminkan apresiasi masyarakat terhadap manfaat program ini, khususnya pada aspek kesehatan dan pendidikan anak. Namun, sentimen negatif menunjukkan kritik terhadap implementasi program, termasuk alokasi anggaran dan dampaknya yang dinilai terlambat untuk isu seperti stunting. Secara keseluruhan, penelitian ini memberikan wawasan penting tentang persepsi publik dan efektivitas pendekatan analisis sentimen dalam mengevaluasi program sosial.

REFERENCES

- [1] R. Saputra and F. N. Hasan, "Analisis Sentimen Terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 3, pp. 411–419, 2024, doi: 10.47233/jteksis.v6i3.1378.
- [2] Z. Purwanti, "Pemodelan Text Mining untuk Analisis Sentimen Terhadap Program Makan Siang Gratis di Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM)," *JIMIK (Jurnal Indones. Manaj. Inform. dan Komunikasi)*, vol. 5, no. 3, pp. 3065–3079, 2024.
- [3] M. J. Abdurrahman and A. Wibowo, "Penerapan Text Mining Menggunakan Metode Multinomial Naive Bayes Program Makan Siang Gratis Pada X," *SENAFTI (Seminar Nas. Mhs. Fak. Teknol. Inf.)*, vol. 3, no. September, pp. 603–611, 2024.
- [4] P. Ardelia Maharani, A. Riyani Namira, and T. Viony Chairunnisa, "Peran Makan Siang Gratis Dalam Janji Kampanye Prabowo Gibran Dan Realisasinya," *Jolasos J. Law Soc. Soc.*, pp. 1–10, 2024.
- [5] F. N. Zaman, M. A. Fadhillah, M. A. Ulinuha, and K. Umam, "Menganalisis Respons Netizen Twitter Terhadap Program Makan Siang Gratis Menerapkan Nlp Metode Naive Bayes," *J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 14, no. 3, pp. 150–233, 2024, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [6] R. A. Hasibuan, D. E. Ratnawati, and R. S. Perdana, "Analisis Sentimen Kebijakan Ekspor Pasir Laut pada Sosial Media Twitter Menggunakan Algoritma Support Vector Machine," *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Inf.*, vol. 5, no. 1, pp. 24–33, 2024, doi: 10.25126/justsi.v5i1.373.
- [7] L. Nursingghah, R. Ruuhwan, and T. Mufizar, "Analisis Sentimen Pengguna Aplikasi X Terhadap Program Makan Siang Gratis Dengan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4336.
- [8] A. Sitanggang, Y. Umaidah, Y. Umaidah, R. I. Adam, and R. I. Adam, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naive Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4902.
- [9] J. M. Br Sembiring and H. H., "Naive Bayes Algorithm Classification in Sentiment Analysis Covid-19 Wikipedia," *J. Tek. Inform.*, vol. 3, no. 4, pp. 869–875, 2022, doi: 10.20884/1.jutif.2022.3.4.311.
- [10] T. Fadiyah Basar, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan Shopeepay dengan Algoritma Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 3, pp. 1426–1433, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] M. R. U. Pulungan, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Ulasan Aplikasi PeduliLindungi dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komun.*, vol. 6, no. 9, pp. 4378–4385, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/11582/5142>
- [12] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305–4313, 2022.
- [13] M. Y. Aldean, P. Paradise, and N. A. Setya Nugraha, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac)," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 2, pp. 64–72, 2022, doi: 10.20895/inista.v4i2.575.



- [14] I. Afdhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, “Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 1, pp. 122–130, 2022, [Online]. Available: <http://ojs.serambimekkah.ac.id/jnkti/article/view/4004/pdf>
- [15] D. S. Ningsih and R. R. Suryono, “Comparison of Naïve Bayes and Information Gain Algorithms in Cyberbullying Sentiment Analysis on Twitter Perbandingan Algoritma Naïve Bayes Dan Information Gain,” *JUTIF (Jurnal Tek. Inform.*, vol. 5, no. 4, pp. 1085–1091, 2024.
- [16] S. Surya Prabu Al Amin, J. Haerul Jaman, and G. Garno, “Analisis Sentimen Masyarakat Terhadap Penanganan Kasus Penembakan Brigadir J Dengan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2519–2526, 2024, doi: 10.36040/jati.v7i4.7126.
- [17] M. J. Palepa, N. Pratiwi, and R. Q. Rohmansa, “Analisis Sentimen Masyarakat Tentang Pengaruh Politik Identitas Pada Pemilu 2024 Terhadap Toleransi Beragama Menggunakan Metode K - Nearest Neighbor,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 1, pp. 389–401, 2024, doi: 10.29100/jupi.v9i1.4957.
- [18] R. Sanjaya, E. Tohidi, E. Wahyudi, and K. Kaslani, “Analisis Sentimen Terhadap Berhentinya Tiktokshop Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 507–514, 2024, doi: 10.36040/jati.v8i1.8443.
- [19] H. C. Husada and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *Teknika*, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
- [20] S. Alam, M. G. Resmi, and N. Masripah, “Classification of Covid-19 vaccine data screening with Naïve Bayes algorithm using Knowledge Discovery in database method,” *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 4, no. 2, pp. 177–185, 2022, doi: 10.47709/cnahpc.v4i2.1584.
- [21] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [22] T. Ridwansyah, “Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [23] D. Pramtanto and F. F. D. Imaniawan, “Analisis Sentimen Terhadap Aplikasi Canva Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbors,” *Comput. Sci.*, vol. 3, no. 2, pp. 110–117, 2023, doi: 10.31294/coscience.v3i2.1917.
- [24] A. Hendra and F. Fitriyani, “Analisis Sentimen Review Halodoc Menggunakan Naïve Bayes Classifier,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 6, no. 2, pp. 78–89, 2021, doi: 10.14421/jjska.2021.6.2.78-89.
- [25] I. Alamsyah and R. T. Shita, “Penerapan Metode Multinomial Naïve Bayes Untuk Implementation of the Multinomial Naïve Bayes Method To Analyze Sentiment of User Reviews of Wahyoo,” *SENAFTI (Prosiding Semin. Nas. Mhs. Fak. Teknol. Inf.)*, vol. 2, no. September, pp. 436–444, 2023.
- [26] Y. A. Singgalen, “Analisis Sentimen dan Sistem Pendukung Keputusan Menginap di Hotel Menggunakan Metode CRISP-DM dan SAW,” *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1343–1353, 2023, doi: 10.47065/josh.v4i4.3917.
- [27] M. H. Wicaksono, M. D. Purbolaksono, and S. Al Faraby, “Perbandingan Algoritma Machine Learning untuk Analisis Sentimen Berbasis Aspek pada Review Female Daily,” *eProceedings Eng.*, vol. 10, no. 3, pp. 3591–3600, 2023.
- [28] H. Utami, “Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network,” *Indones. J. Appl. Stat.*, vol. 5, no. 1, p. 31, 2022, doi: 10.13057/ijas.v5i1.56825.
- [29] G. A. Ramadhan, *Analisis sentimen ulasan aplikasi ruangguru dengan algoritma long short term memory*. 2023. [Online]. Available: [https://repository.uinjkt.ac.id/dspace/handle/123456789/71510%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/71510/1/GILANG AMBANG RAMADHAN-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/71510%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/71510/1/GILANG%20AMBANG%20RAMADHAN-FST.pdf)
- [30] R. Faiz Ananda, A. Syahri, and F. N. Hasan, “Sentiment Analysis of Customer Satisfaction in Gojek and Grab Application Reviews Using the Naïve Bayes Algorithm,” *J. Tek. Inform. (JUTIF).I*, vol. 5, no. 1, pp. 233–241, 1680, [Online]. Available: <https://doi.org/10.52436/1.jutif.2024.5.1.1680>
- [31] M. R. A. Nasution and M. Hayaty, “Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter,” *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.