PENERAPAN NAÏVE BAYES UNTUK ANALISIS SENTIMEN NETIZEN TERHADAP PROGRAM MAKAN SIANG GRATIS PADA MEDIA SOCIAL X

Ririn Amelia Br Siregar, Albert Ramadhan Manik, Alfin Syahri, Muhammad Budi Akbar, Arnita, Fanny Ramadhani
Ilmu Komputer, Universitas Negeri Medan
Jalan W. Iskandar Psr V Medan Esatate Kab. Deli Serdang , Indonesia
rierienamelia.4231250003@mhs.unimed.ac.id

ABSTRAK

Program Makan Siang Gratis yang diperkenalkan dalam Pemilihan Presiden 2024 menjadi perhatian publik dan menuai berbagai tanggapan, baik positif maupun negatif. Persepsi masyarakat terhadap program ini penting untuk diketahui guna mengevaluasi keberhasilan implementasi kebijakan. Penelitian ini bertujuan untuk menganalisis sentimen netizen di media sosial X terhadap program tersebut menggunakan algoritma Naïve Bayes. Data dikumpulkan melalui proses crawling dengan Twitter API, kemudian dilakukan preprocessing, ekstraksi fitur menggunakan TF-IDF, dan klasifikasi sentimen ke dalam tiga kategori: positif, negatif, dan netral. Hasil evaluasi menunjukkan bahwa model Naïve Bayes yang digunakan memiliki akurasi sebesar 65%, presisi 71%, recall 65%, dan F1-score 62%. Analisis menunjukkan bahwa sentimen negatif mendominasi, disusul oleh sentimen netral, sementara sentimen positif hanya sedikit. Temuan ini mengindikasikan adanya kekhawatiran masyarakat terhadap efektivitas dan transparansi program. Hasil penelitian ini dapat menjadi bahan pertimbangan bagi pemerintah dalam meningkatkan sosialisasi dan implementasi program agar lebih diterima oleh masyarakat..

Kata kunci : *Analisis Sentimen, Naïve Bayes, Media Sosial, Program Makan Siang Gratis, Twitter.*

1. PENDAHULUAN

Salah satu inisiatif yang diperkenalkan oleh pasangan calon Prabowo Subianto dan Gibran Rakabuming Raka dalam Pemilihan Presiden 2024 adalah Program Makan Siang Gratis. Program ini menarik perhatian publik karena menargetkan kelompok rentan seperti siswa sekolah, santri, ibu hamil, dan balita. Tujuan utama dari program ini adalah untuk mengurangi angka kemiskinan, mencegah stunting, serta mengatasi permasalahan gizi buruk pada anak-anak di Indonesia. Namun, keberhasilan sebuah program pemerintah tidak hanya ditentukan oleh efektivitas distribusi bantuan atau pengelolaan anggaran, tetapi juga oleh bagaimana masyarakat memandang program tersebut, karena persepsi publik memainkan peran penting dalam keberlanjutan serta kesuksesan kebijakan di masa mendatang.[1]

Stunting, sebagai dampak dari kurang gizi kronis, masih menjadi masalah serius di Indonesia. Data terbaru menunjukkan bahwa prevalensi stunting di Indonesia mencapai 21,6% pada tahun 2022. Program makan siang gratis diharapkan dapat menjadi solusi untuk mengatasi masalah ini dengan memberikan akses makanan bergizi kepada anak-anak di sekolah dan pesantren. Namun, seperti halnya kebijakan publik lainnya, program ini menuai berbagai tanggapan dari masyarakat, mulai dari dukungan hingga kritik.

Pembahasan mengenai program tersebut mendapatkan berbagai tanggapan dari netizen, salah satunya berasal dari pengguna platform X. Platform-platform online memainkan peran yang sangat penting dalam menjangkau dan menyebarkan informasi tentang berbagai program sosial. Selain itu, platform-platform ini juga menjadi tempat utama bagi diskusi publik dan memberikan kesempatan bagi pengguna untuk menyampaikan pendapat mereka tentang program tersebut. Banyak pendapat yang muncul mengenai program kerja yang diajukan oleh Presiden Prabowo Subianto dan Wakil Presiden Gibran Rakabuming Raka, termasuk program makan siang gratis.[2]

Dengan menawarkan makan siang gratis, prabowo-gibran berusaha memberikan perlindungan sosial kepada mereka yang rentan terhadap kemiskinan dan ketidaksetaraan, serta menjamin bahwa semua orang dapat berpartisipasi dalam kemajuan bangsa . Salah satu program bantuan gizi yang memberikan makan dan susu gratis kepada anak balita dan ibu hamil disekolah dah pesantren. Anak di prasekolah, sekolah dasar, sekolah menengah pertama, sekolah menengah atas, dan pesantren mendapatkan makan siang gratis setiap hari. Program makan siang gratis memiliki tujuan yaitu dikategorikan ke dalam indikator pencapaian. Berbagi susu gratis kepada ibu hamil lebih relevan sebagai penanganan stunting daripada upaya untuk mengatasi melalui program ini. Program makan siang gratis tampaknya memiliki dampak positif di banyak negara. Makan siang gratis ini meningkatkan nutrisi anak, partisipasi sekolah, karakter siswa, dan ekonomi keluarga di seluruh negeri.[3]

Platform X adalah salah satu media sosial yang berkembang pesat, dimana penggunanya dapat menghasilkan, memposting, serta membaca pesan teks pendek atau yang biasa dianggap sebagai tweet. Melaluitweet tersebut pengguna platform X diberikan kebebasan untuk mengungkapkan pendapat dan pernyataan

mengenai layanan, institusi, atau topik politik atau perdebatan.[4]

Penelitian ini bertujuan untuk menerapkan algoritma Naïve Bayes dalam menganalisis sentimen netizen di media sosial X terhadap program makan siang gratis. Dengan menggunakan data dari X, penelitian ini akan mengidentifikasi dan mengklasifikasikan sentimen masyarakat terhadap program tersebut. Hasil analisis ini diharapkan dapat memberikan wawasan yang mendalam tentang respons publik terhadap program makan siang gratis, serta menjadi bahan pertimbangan bagi pemerintah dalam merumuskan kebijakan yang lebih efektif dan diterima oleh masyarakat.[5]

Penelitian ini bertujuan untuk menerapkan algoritma Naïve Bayes dalam menganalisis sentimen netizen di media sosial X terhadap program makan siang gratis. Dengan menggunakan data dari X, penelitian ini akan mengidentifikasi dan mengklasifikasikan sentimen masyarakat terhadap program tersebut. Hasil analisis ini diharapkan dapat memberikan wawasan yang mendalam tentang respons publik terhadap program makan siang gratis, serta menjadi bahan pertimbangan bagi pemerintah dalam merumuskan kebijakan yang lebih efektif dan diterima oleh masyarakat.

2. TINJAUAN PUSTAKA

mengeksplorasi penerapan algoritma Naïve Bayes dalam menganalisis sentimen ulasan pengguna pada platform e-commerce. Studi ini menekankan pentingnya algoritma Naïve Bayes dalam mengklasifikasikan sentimen positif, negatif, dan netral dari ulasan pengguna. Hasil penelitian menunjukkan bahwa algoritma ini efektif dalam memproses data teks dalam jumlah besar, memberikan akurasi yang signifikan dalam klasifikasi sentimen, dan dapat membantu platform e-commerce dalam memahami persepsi dan kepuasan pelanggan.[6]

menerapkan algoritma Naïve Bayes untuk menganalisis sentimen pengguna terhadap aplikasi kesehatan digital, khususnya Halodoc. Dengan membagi data ulasan pengguna ke dalam kategori positif, negatif, dan netral, penelitian ini menemukan bahwa algoritma Naïve Bayes mampu mencapai tingkat akurasi hingga 86% dalam klasifikasi sentimen. Temuan ini menegaskan efektivitas algoritma Naïve Bayes dalam menganalisis sentimen pengguna pada aplikasi kesehatan digital, memberikan wawasan berharga bagi pengembang aplikasi dalam meningkatkan layanan mereka.[7]

Dalam penelitian ini, algoritma Naive Bayes dipilih karena memiliki beberapa keunggulan yang mendukung efektivitas analisis sentimen teks. Yang pertama adalah bahwa algoritma ini dapat memproses jumlah besar teks dengan kompleksitas komputasi yang rendah, yang membuatnya sangat cocok untuk mengolah dataset media sosial yang biasanya sangat besar. Kedua, meskipun algoritma ini cukup sederhana untuk digunakan, ia masih efektif dalam mengklasifikasikan teks menjadi kategori tertentu, seperti positif, negatif, atau Hasil penelitian sebelumnya menunjukkan bahwa algoritma ini memiliki kinerja yang kompetitif dalam berbagai konteks analisis sentimen.

Sebagai contoh, Saputra & Hasan (2024) menemukan bahwa menggunakan Naive Bayes untuk analisis sentimen pada program serupa mencapai akurasi 65%, presisi 71%, recall 65%, dan skor F1 62%. Selain itu, Setiawan et al. (2024) menggunakan algoritma yang sama untuk mencatat akurasi hingga 86% dalam analisis sentimen aplikasi kesehatan digital. Temuan ini menunjukkan bahwa Naive Bayes dapat memberikan hasil yang dapat diandalkan dan bersaing untuk analisis sentimen teks pada media sosial, termasuk dalam Program Makan Siang Gratis. Penelitian ini diharapkan dapat memberikan wawasan mendalam tentang respons publik terhadap program serta menjadi bahan pertimbangan bagi pemerintah untuk meningkatkan sosialisasi dan pelaksanaan program agar masyarakat lebih menerimanya.

2.1. Analisis Sentimen

Era media sosial sudah dekat. Orang cenderung menggunakan alat media sosial seperti Twitter untuk menyiarkan suasana hati, pendapat, dan status mereka. Sentimen orang dapat diukur melalui analisis sentimen Twitter.[8]

Analisis sentimen adalah langkah evaluasi dan pengukuran perasaan, pandangan, atau respon dari individu atau kelompok terhadap suatu topik atau entitas. Proses ini umumnya dilakukan dengan menggunakan teknik komputasional atau metode analisis teks untuk memutuskan apakah sentimen tersebut cenderung positif, negatif, atau netral. Analisis sentimen, yang kadang disebut sebagai opinion mining, mengindikasikan bahwa terdapat emosi yang tersirat di dalam kata-kata yang dipakai oleh individu.[2]

2.2. Algoritma Naïve Bayes

Mengeksplorasi penerapan Algoritma Naïve Bayes untuk mendeteksi hoaks di media sosial. Algoritma ini dipilih karena kemampuannya dalam memproses data teks berskala besar dengan kompleksitas yang relatif rendah dan implementasi yang mudah. Dataset yang digunakan terdiri dari postingan media sosial yang dikategorikan sebagai hoaks atau bukan hoaks. Proses pra-pemrosesan data mencakup tokenisasi, ekstraksi fitur, dan pembersihan teks menggunakan teknik TF-IDF (Term Frequency-Inverse Document Frequency). Hasil penelitian menunjukkan bahwa Algoritma Naïve Bayes dapat mengidentifikasi hoaks dengan tingkat kesalahan yang sangat rendah, menjadikannya efektif dalam mendeteksi konten hoaks di media sosial.[9]

menerapkan Algoritma Naïve Bayes untuk memprediksi kelulusan siswa. Dalam penelitian ini, data kelulusan siswa diproses menggunakan teknik data mining dengan Algoritma Naïve Bayes untuk

memprediksi kelulusan siswa. Atribut seperti Nilai Praktik, Ujian Sekolah (US), Ujian Nasional (UN), dan Perilaku Siswa digunakan dalam proses prediksi ini. Hasil penelitian menunjukkan bahwa dari populasi 60 siswa, model berhasil memprediksi bahwa 45 siswa akan lulus, dan 15 siswa tidak lulus. Algoritma Naïve Bayes yang diimplementasikan menggunakan aplikasi Orange menghasilkan tingkat akurasi sebesar 98,33% dengan tingkat presisi 100,00%. Hasil ini menunjukkan bahwa Algoritma Naïve Bayes sangat efektif dalam memprediksi kelulusan siswa dengan tingkat akurasi yang tinggi.[10]

2.3. Program Makan Siang Gratis

Penelitian yang dilakukan oleh Asro dan Sulaiman (2024) memanfaatkan platform YouTube untuk mempelajari tanggapan publik terhadap perubahan dari program makan siang gratis menjadi program makan siang yang sehat. Studi ini menemukan dengan menggunakan regresi logistik bahwa mayoritas komentar pengguna bersifat positif, menunjukkan dukungan publik terhadap program gizi yang lebih baik. Temuan ini menekankan betapa pentingnya berkomunikasi dengan baik untuk menyebarkan program pemerintah.[11]

Saputra dan Hasan (2024) menggunakan algoritma Naive Bayes untuk menganalisis sentimen publik terhadap program makan siang dan susu gratis. Hasil penelitian menunjukkan bahwa sentimen positif mendominasi, mencerminkan apresiasi masyarakat terhadap upaya pemerintah dalam meningkatkan gizi anak-anak. Studi ini menyoroti efektivitas algoritma Naive Bayes dalam mengkategorikan sentimen publik.[5]

Fasha dan Tesniyadi menganalisis artikel di Tempo.co yang membahas penggunaan Dana BOS untuk program makan siang gratis. Menurut penelitian ini, ada perdebatan tentang bagaimana dana tersebut dialokasikan. Beberapa orang khawatir bahwa menggunakan dana BOS untuk program makan siang dapat mengurangi anggaran untuk tujuan pendidikan lainnya. Dalam implementasi kebijakan publik, transparansi dan komunikasi yang jelas sangat penting, menurut penelitian ini.[12]

Penelitia Ulfatul karomah dkk menyelidiki hubungan antara program makan siang gratis dan tingkat kehadiran dan prestasi akademik siswa di sekolah dasar. Hasilnya menunjukkan bahwa program makan siang gratis meningkatkan kehadiran dan nilai akademik siswa, terutama dalam mata pelajaran matematika dan sains. Hal ini menunjukkan bahwa program makan siang dapat membantu siswa belajar dengan lebih baik.[13]

2.4. Studi Terkait

Penelitian yang dilakukan oleh Merlinda dan Yusuf (2025) membahas dampak Program Makan Bergizi Gratis (MBG) terhadap motivasi belajar siswa dari sudut pandang sosiologi pendidikan. Penelitian ini terkait dengan proposal penelitian ini. Studi ini menunjukkan bahwa program makan siang gratis memiliki potensi besar untuk meningkatkan kualitas makanan siswa dan konsentrasi mereka di sekolah, yang secara langsung berkontribusi pada prestasi akademik mereka. Penelitian ini menekankan bagaimana implementasi program ini dapat dipengaruhi oleh kebijakan publik dan persepsi masyarakat terhadapnya, menggunakan analisis wacana kritis Teun A. Van Dijk. Selain itu, hasil penelitian ini mendukung hipotesis bahwa memberi siswa makanan bergizi gratis tidak hanya mengurangi jumlah orang yang kelaparan di sekolah tetapi juga membantu mereka mulai menerapkan kebiasaan makan yang sehat sejak kecil.Penelitian ini dapat menjadi referensi penting dalam menganalisis bagaimana algoritma Naïve Bayes dapat diterapkan untuk memahami persepsi netizen terhadap program ini di media sosial X, sebagaimana direncanakan dalam penelitian yang diusulkan.[14]

Berdasarkan studi terkait, penelitian oleh Ikhsan et al. (2023) dalam jurnal Infotekmesin menganalisis sentimen publik terhadap kebijakan kenaikan Pajak Pertambahan Nilai (PPN) menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM)

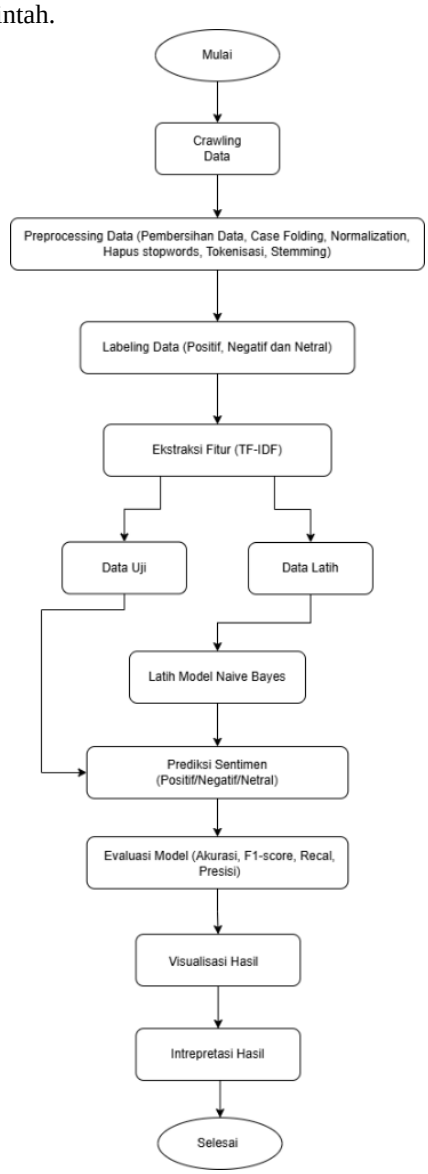
Studi ini menunjukkan bahwa media sosial, khususnya platform X, menjadi wadah utama bagi masyarakat dalam menyuarakan pendapatnya terhadap kebijakan pemerintah. Relevansi penelitian ini dengan proposal yang diajukan terletak pada metode yang digunakan, yaitu analisis sentimen berbasis algoritma Naïve Bayes, yang juga diterapkan dalam penelitian tentang persepsi netizen terhadap Program Makan Siang Gratis. Dengan membandingkan hasil akurasi kedua algoritma sebelum dan sesudah proses balancing data, penelitian ini memberikan gambaran tentang efektivitas Naïve Bayes dalam mengolah data opini publik. Temuan tersebut dapat menjadi acuan dalam penelitian ini untuk memahami bagaimana sentimen publik terbentuk, serta bagaimana data media sosial dapat digunakan sebagai indikator dalam evaluasi kebijakan pemrintah[15]

3. METODE PENELITIAN

Proses pengumpulan data dilakukan dalam beberapa tahap. Pertama, *crawling* data dilakukan secara otomatis menggunakan Twitter API dengan bantuan library Tweepy dalam bahasa pemrograman Python. Setelah data terkumpul, dilakukan *filtering* untuk menghapus tweet yang tidak relevan, seperti spam atau promosi, guna memastikan hanya opini yang benar-benar membahas Program Makan Siang Gratis yang digunakan. Selanjutnya, data melalui tahap *preprocessing*, yang mencakup penghapusan karakter khusus, angka, tanda baca, *stopwords*, URL, dan emoji, *Normalization*, *Tokenizing* serta proses *stemming* untuk mengonversi teks ke dalam bentuk kata dasar agar meningkatkan akurasi analisis sentimen.

Tahapan analisis data dalam penelitian ini diawali dengan *preprocessing* untuk membersihkan

data dan memastikan bahwa hanya kata-kata bermakna yang digunakan dalam analisis. Setelah itu, dilakukan klasifikasi sentimen menggunakan algoritma Naïve Bayes, di mana model ini dilatih dengan dataset latih sebelum diterapkan pada data hasil *crawling*. Proses klasifikasi bertujuan untuk mengelompokkan opini publik ke dalam tiga kategori sentimen, yaitu positif, negatif, atau netral. Setelah klasifikasi dilakukan, hasil analisis sentimen disajikan dalam beberapa bentuk, seperti visualisasi data menggunakan diagram batang atau *pie chart* untuk menunjukkan distribusi sentimen netizen, serta tabel yang menampilkan contoh tweet dari masing-masing kategori sentimen. Interpretasi hasil juga dilakukan untuk mengidentifikasi faktor-faktor yang memengaruhi sentimen masyarakat terhadap Program Makan Siang Gratis, sehingga dapat digunakan sebagai bahan evaluasi kebijakan pemerintah.



Gambar 1. Flowchart

Penelitian ini menggunakan berbagai alat dan perangkat lunak untuk mendukung pengumpulan, analisis, dan penyajian data. Python digunakan sebagai bahasa pemrograman utama, dengan beberapa library pendukung seperti Tweepy untuk mengambil data dari Twitter API, Pandas untuk manipulasi dan analisis data, serta NLTK untuk *preprocessing* teks. Implementasi algoritma Naïve Bayes dan evaluasi model dilakukan menggunakan Scikit-learn, sementara Google Colab dimanfaatkan sebagai platform pemrosesan data. Untuk keperluan visualisasi, Matplotlib dan Seaborn digunakan dalam menghasilkan grafik yang lebih informatif dan mudah dipahami.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh melalui proses crawling menggunakan Tweepy di Google Colab. Proses pengambilan data dilakukan dengan menargetkan tweet yang membahas program makan siang gratis. Sumber data berasal dari aplikasi X, dengan pencarian dilakukan menggunakan kata kunci "program makan siang gratis". Hasil crawling menghasilkan sebanyak 3.042 tweet yang dikumpulkan dalam rentang waktu dari Desember 2024 hingga Februari 2025. Data ini mencakup berbagai opini netizen terkait program tersebut, baik dalam sentimen positif, negatif, maupun netral.

Data yang diperoleh disimpan dalam bentuk file CSV yang berisi beberapa kolom, di antaranya created_at, id_str, full_text, quote_count, reply_count, retweet_count, favorite_count, lang, user_id_str, conversation_id_str, username, dan tweet_url**.** Namun, dalam penelitian ini, hanya kolom full_text yang digunakan, karena berisi isi lengkap dari setiap tweet yang akan dianalisis sentimennya.

4.2. Preprocessing Data

Sebelum data dianalisis, beberapa tahap preprocessing dilakukan untuk membersihkan dan mempersiapkan teks agar algoritma Naive Bayes lebih mudah diproses. Tahapan-tahap ini adalah sebagai berikut:

a. Pembersihan Data : Proses pembersihan data tweet akan dilakukan dengan menghilangkan atribut yang tidak diperlukan, seperti URL, hashtag, emotikon, dan mention.

Tabel 1. Proses Pembersihan Data

Sebelum	Sesudah
@stackedsores @prabowo Bapak @prabowo jangan malu-malu gitu lah masa gengsinya sebatas program makan siang gratis sih.	Bapak jangan malumalu gitu lah masa gengsinya sebatas program makan siang gratis sih
@zer0man1407 @watermelayen @Agakgini bisa ga sih? secara mereka gaada program buang buang duit buat makan siang sama susu gratis. pendidikan gratis mah oke buat semua orang bro	bisa ga sih secara mereka gaada program buang buang duit buat makan siang sama susu gratis pendidikan gratis mah oke buat semua orang bro

b. Case Folding : Proses mengubah seluruh teks menjadi huruf kecil agar konsistensi data tetap terjaga dan tidak terjadi perbedaan makna akibat perbedaan huruf besar dan kecil.

Tabel 2. Proses Case Folding

Sebelum	Sesudah
LAH KOCAK PROGRAM MAKAN SIANG GRATIS NGORBANIN FASILITAS UMUM	lah kocak program makan siang gratis ngorbanin fasilitas umum
Demi sebuah program makan siang yang katanya GRATIS	demi sebuah program makan siang yang katanya gratis

c. Normalization : Proses mengubah kata-kata tidak baku atau singkatan menjadi bentuk yang lebih standar.

Tabel 3. Proses Normalization

Sebelum	Sesudah
program ga jelas anak makan siang gratis seluruh anggota keluarga bs jd kena imbasnya	program tidak jelas anak makan siang gratis seluruh anggota keluarga bisa jd kena imbasnya
inilah program makan siang gratis yg gw mau	inilah program makan siang gratis yang saya ingin

d. Hapus Stopword : Menghapus kata-kata umum (stop words) seperti "dan", "di", "yang", dan sejenisnya yang tidak memiliki makna signifikan dalam analisis sentimen.

Tabel 4. Proses Hapus Stopword

Sebelum	Sesudah
gizi untuk generasi bangsa kesehatan yang baik mendukung pendidikan yang optimal bagus sekali program makan siang gratis ini	gizi generasi bangsa kesehatan baik mendukung pendidikan optimal bagus sekali program makan siang gratis

e. Tokenize : Proses memecah teks menjadi kata-kata atau token agar lebih mudah dianalisis.

Tabel 5. Proses Tokenize

Sebelum	Sesudah
demi sebuah program makan siang katanya gratis	demi, sebuah, program, makan, siang, katanya, gratis

f. Stemming : Proses mengubah kata-kata menjadi bentuk dasar atau akar katanya, misalnya "makanannya" menjadi "makan" agar analisis lebih akurat.

Tabel 6. Proses Stemming

Sebelum	Sesudah
bapak jangan malu malu gitu lah masa gengsinya sebatas program makan siang gratis sih	bapak jangan malu malu gitu masa gengsi batas program makan siang gratis

4.3. Labeling Sentimen

Setelah tahap preprocessing selesai, data kemudian diberi label sentimen untuk menentukan kategori opini dalam setiap tweet. Proses pelabelan dimulai dengan menentukan apakah setiap tweet menunjukkan sentimen positif, netral, atau negatif terhadap program tersebut. Sentimen positif menunjukkan dukungan atau pandangan baik terhadap program makan siang gratis, sentimen negatif mencerminkan ketidaksetujuan atau kritik, sementara sentimen netral berisi opini yang tidak condong ke arah positif maupun negatif.

Kami melakukan proses labeling menggunakan library IndoBERT di Google Colab untuk mengklasifikasikan sentimen dalam setiap tweet. Setelah itu, kami mengecek ulang hasil labeling guna memastikan kesesuaiannya. Jika ditemukan ketidaksesuaian dalam klasifikasi sentimen, kami melakukan koreksi secara manual agar data yang digunakan lebih akurat.

Selanjutnya, data dibagi menjadi dua bagian, yaitu data latih dan data uji dengan rasio 80:20 untuk keperluan pelatihan dan pengujian model analisis sentimen.

Tabel 7. Proses Labeling

Komentar	Sentimen
makan siang gratis program paling gak masuk akal semua kena imbasnya sekarang	Negatif
aku petugas program makan siang gratis	Netral
ayo dukung kesehatan yang baik mendukung pendidikan yang optimal bagus sekali program makan siang gratis ini	Positif

Hasil pelabelan ini kemudian digunakan sebagai data latih untuk proses pembuatan model analisis sentiment

4.4. Ekstraksi Fitur

Setelah tahap labeling data selesai, langkah selanjutnya adalah ekstraksi fitur menggunakan metode TF-IDF (Term Frequency - Inverse Document Frequency). TF-IDF digunakan untuk menemukan kata-kata yang paling penting dalam dokumen atau kumpulan dokumen. Konversi teks menjadi vektor fitur numerik adalah bagian dari ekstraksi fitur dalam analisis teks. Proses ini penting dalam mempersiapkan data untuk analisis lanjutan seperti klasifikasi atau prediksi.

Tabel 8. Nilai Fitur TF-IDF

Kata	Total TF-IDF
Makan	190.757662
Program	183.528795
Gratis	181.909392
Siang	174.301981
Tidak	116.513817

Berdasarkan Tabel 8, nilai fitur TF-IDF menunjukkan tingkat kepentingan kata-kata dalam dokumen. Kata "Makan" memiliki nilai TF-IDF tertinggi sebesar 190.757662, menunjukkan bahwa

kata ini lebih signifikan dalam dokumen dibandingkan kata lainnya. Kata "Program" dan "Gratis" juga memiliki nilai TF-IDF yang cukup tinggi, masing-masing sebesar 183.528795 dan 181.909392, menandakan bahwa kata-kata ini sering muncul dalam dokumen tertentu tetapi tidak terlalu umum di seluruh dokumen. Sebaliknya, kata "Tidak" memiliki nilai TF-IDF terendah sebesar 116.513817, yang menunjukkan bahwa kata ini mungkin lebih sering muncul di berbagai dokumen dan kurang memiliki bobot dalam membedakan suatu dokumen dari yang lain. Dengan demikian, analisis TF-IDF membantu dalam mengidentifikasi kata-kata yang lebih relevan dan berpengaruh dalam suatu teks, sehingga dapat digunakan dalam berbagai aplikasi seperti pencarian informasi dan klasifikasi dokumen.

4.5. Pembagian Data Training dan Data Uji

Setelah tahap ekstraksi fitur selesai, data dibagi menjadi data latihan (training data) dan data uji (testing data). Data latihan digunakan untuk membuat model klasifikasi sentimen, dan data uji digunakan untuk mengevaluasi kinerja model dalam mengklasifikasikan tweet baru.

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 mereka dapat secara objektif mengevaluasi akurasi dan kinerja model sebelum menerapkannya pada data baru.

4.6. Latih Model Naïve Bayes

```
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, classification_report

# Inisialisasi model Naïve Bayes
nb_model = MultinomialNB()

# Latih model dengan data training
nb_model.fit(X_train, y_train)

# Prediksi data uji
y_pred = nb_model.predict(X_test)
```

Gambar 2. Program Latih Model Naïve Bayes

Salah satu metode pembelajaran probabilistik yang banyak digunakan dalam pengolahan bahasa alami (Natural Language Processing/NLP) adalah algoritma Multinomial Naïve Bayes, yang didasarkan pada Teorema Bayes. Algoritma ini bekerja dengan menghitung probabilitas suatu kelas berdasarkan frekuensi kemunculan kata dalam sebuah dokumen. Dalam konteks klasifikasi teks, Multinomial Naïve Bayes sangat efektif dalam menangani representasi fitur berbasis Term Frequency-Inverse Document Frequency (TF-IDF). Pada tahap ini, model Naïve Bayes dilatih menggunakan data latih yang telah diproses sebelumnya untuk membangun model klasifikasi sentimen. Multinomial Naïve Bayes dipilih karena kesederhanaannya serta kemampuannya dalam menangani data teks yang memiliki distribusi multinomial, menjadikannya salah satu metode yang paling umum digunakan dalam klasifikasi teks.

Proses pelatihan model Naïve Bayes terdiri dari beberapa tahap utama. Pertama, dilakukan persiapan data, yaitu mengubah teks mentah menjadi representasi numerik menggunakan teknik TF-IDF. Langkah ini bertujuan untuk mengubah kata-kata menjadi vektor angka sehingga dapat diolah oleh algoritma. Kedua, data dibagi menjadi dua bagian, yaitu data latih dan data uji, di mana data latih digunakan untuk membangun model sementara data uji digunakan untuk mengukur kinerjanya. Ketiga, model Multinomial Naïve Bayes diinisialisasi menggunakan MultinomialNB() dari pustaka scikit-learn, kemudian dilatih dengan memanggil fungsi fit(X_train, y_train), di mana X_train merupakan fitur hasil transformasi teks, dan y_train adalah label kelas sentimen.

Setelah model dilatih, tahap selanjutnya adalah melakukan prediksi sentimen terhadap data uji. Model yang telah dipelajari akan menganalisis fitur-fitur dari data uji dan mengklasifikasikan setiap teks ke dalam kategori sentimen yang telah ditentukan sebelumnya, seperti positif, negatif, atau netral, menggunakan perintah predict(X_test). Hasil prediksi ini kemudian dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score untuk mengukur seberapa baik model dalam melakukan klasifikasi. Dengan pendekatan ini, sistem dapat secara otomatis mengelompokkan teks berdasarkan sentimen yang terkandung di dalamnya, sehingga mempermudah analisis dalam berbagai aplikasi NLP.

4.7. Evaluasi

Setelah prediksi dilakukan, hasil prediksi sentimen yang dibuat menggunakan algoritma Naive Bayes dievaluasi. Ini dilakukan dengan metrik seperti akurasi, presisi, recall, dan skor F1. Hasilnya menunjukkan seberapa baik model mampu mengklasifikasikan sentimen dengan benar.

```
Laporan Klasifikasi:
              precision    recall  f1-score   support

     Negatif       0.61      0.93      0.74       274
      Netral       0.73      0.50      0.59       268
     Positif       1.00      0.12      0.22        64

    accuracy                           0.65       606
   macro avg       0.78      0.52      0.52       606
weighted avg       0.71      0.65      0.62       606
```

Gambar 3. Hasil Evaluasi

Confusion matrix digunakan untuk menganalisis jumlah prediksi yang benar dan salah dalam setiap kategori sentimen (negatif, netral, dan positif). Dari laporan klasifikasi yang dihasilkan dapat dilihat pada tabel .

Tabel 9. Hasil Metrik Evaluasi

Akurasi	65%
Presisi	71%
Recall	65%
F1-Score	62%

Model Naive Bayes memiliki akurasi 65%, presisi 71%, recall 65%, dan skor F1 62%. Hasil ini menunjukkan bahwa model cukup baik dalam mengklasifikasikan sentimen, tetapi masih memiliki kelemahan dalam mengenali kelas tertentu, terutama kategori Positif yang memiliki recall rendah.

Evaluasi model dilakukan menggunakan beberapa metrik utama, yaitu akurasi, presisi, recall, dan F1-Score. Akurasi mengukur seberapa sering model membuat prediksi yang benar dibandingkan dengan keseluruhan data, yang dapat dihitung menggunakan rumus:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Dari hasil evaluasi, model memiliki akurasi 65%, yang berarti model mengklasifikasikan data dengan benar dalam 65% dari semua kasus. Selain itu, presisi digunakan untuk mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan total prediksi positif yang dibuat oleh model, yang dihitung dengan rumus:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

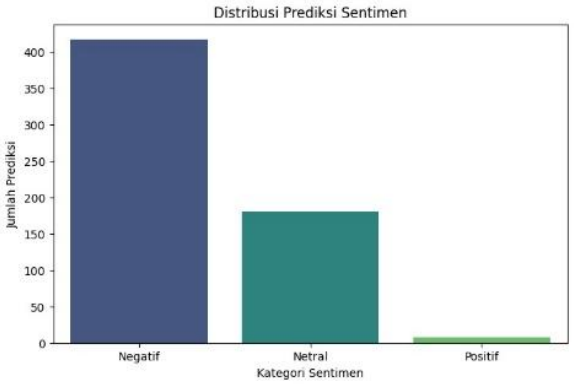
Dari hasil klasifikasi, presisi untuk kelas negatif, netral, dan positif masing-masing adalah 0.61, 0.73, dan 1.00, dengan rata-rata berbobot sebesar 0.71. Nilai presisi yang tinggi pada kelas positif menunjukkan bahwa ketika model mengklasifikasikan suatu sampel sebagai positif, hasil tersebut hampir selalu benar. Namun, nilai ini perlu dikombinasikan dengan recall, yang mengukur seberapa banyak data positif yang benar-benar dikenali oleh model, yang dirumuskan sebagai berikut:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

Dari hasil evaluasi, nilai F1-Score untuk kelas negatif, netral, dan positif masing-masing adalah 0.74, 0.59, dan 0.22, dengan rata-rata berbobot sebesar 0.62. Nilai F1-Score yang rendah pada kelas positif menunjukkan bahwa model masih kesulitan dalam mengenali dan mengklasifikasikan data positif dengan baik.

Secara keseluruhan, model Naive Bayes yang digunakan dalam penelitian ini memiliki performa yang cukup baik dalam mengklasifikasikan sentimen, dengan akurasi sebesar 65%. Namun, kelemahan utama terletak pada kemampuan model dalam mengenali kelas positif, sebagaimana terlihat dari nilai recall yang sangat rendah. Hal ini dapat disebabkan oleh distribusi data yang tidak seimbang, di mana jumlah sampel untuk kategori positif lebih sedikit dibandingkan dengan kategori lainnya. Untuk meningkatkan performa model, beberapa pendekatan yang dapat dilakukan adalah dengan menyeimbangkan data menggunakan teknik oversampling atau undersampling, mencoba model lain seperti Random Forest atau SVM, atau meningkatkan fitur ekstraksi, misalnya dengan menggunakan TF-IDF atau word embeddings agar model dapat menangkap informasi yang lebih baik dari teks.

4.8. Visualisasi Hasil



Gambar 4. Visualisasi Hasil Prediksi

Hasil prediksi sentimen menunjukkan bahwa mayoritas masyarakat memiliki pandangan negatif terhadap program makan siang gratis. Hal ini bisa mengindikasikan adanya kekhawatiran terkait pelaksanaan program, seperti kualitas makanan, efektivitas distribusi, atau transparansi anggaran. Sentimen netral juga cukup signifikan, menunjukkan bahwa sebagian masyarakat masih ragu atau belum memiliki pendapat yang kuat.

Sementara itu, hanya sedikit yang memberikan respons positif, yang mengindikasikan bahwa dukungan terhadap program ini masih terbatas. Temuan ini dapat menjadi bahan evaluasi bagi pemerintah atau penyelenggara untuk meningkatkan kepercayaan publik, misalnya dengan memperbaiki sistem pelaksanaan atau menyosialisasikan manfaat program secara lebih efektif.

5. KESIMPULAN DAN SARAN

Hasil analisis sentimen yang dilakukan terhadap program makan siang gratis di media sosial X menggunakan algoritma Naive Bayes menunjukkan akurasi model sebesar 65%, presisi sebesar 71%, recall sebesar 65%, dan nilai F1 sebesar 62%. Hasil klasifikasi menunjukkan bahwa model lebih akurat dalam mengidentifikasi sentimen negatif daripada netral dan positif, dengan nilai recall tertinggi dalam kategori negatif (0.93) dan nilai recall terendah dalam kategori positif (0.12), yang menunjuk Kemudian, berdasarkan hasil analisis sentimen, dapat disimpulkan bahwa pendapat orang di media sosial tentang program makan siang gratis cenderung didominasi oleh opini negatif. Ada kekhawatiran tentang kinerja program, ketidakjelasan anggaran, atau pelaksanaan yang tidak memuaskan. Sentimen netral menunjukkan bahwa ada sebagian masyarakat yang bersikap objektif atau belum memiliki opini yang kuat terhadap program ini. Sementara itu, jumlah sentimen positif yang rendah mengindikasikan bahwa hanya sedikit pengguna yang secara terbuka menyampaikan dukungan terhadap kebijakan ini.

Hasil penelitian memberikan beberapa rekomendasi untuk penelitian selanjutnya yaitu, menggunakan dataset yang lebih besar dan seimbang serta menerapkan teknik "balancing data" seperti

SMOTE untuk meningkatkan representasi kelas minoritas dan mengoptimalkan preprocessing dengan teknik "embedding word" (Word2Vec, GloVe, dan sebagainya). Diharapkan bahwa tindakan ini akan memperbaiki kelemahan yang ada dan menghasilkan hasil yang lebih akurat dan relevan.

DAFTAR PUSTAKA

[1] Tundo and D. N. Rachmawati, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Terhadap Program Makan Siang Gratis," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 3, pp. 411–419, 2024, doi: 10.47233/jteksis.v6i3.1378.

[2] F. N. Zaman, M. A. Fadhilah, M. A. Ulinuha, and K. Umam, "Menganalisis Respons Netizen Twitter Terhadap Program Makan Siang Gratis Menerapkan Nlp Metode Naïve 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

[3] M. J. Abdurrahman and A. Wibowo, "PENERAPAN TEXT MINING MENGGUNAKAN METODE MULTINOMIAL NAIVE BAYES PROGRAM MAKAN SIANG GRATIS PADA X," vol. 3, no. September, pp. 603–611, 2024.

[4] M. Rafli Tanjung, N. Heryana, and S. Siska, "Analisis Sentimen Pengguna Platform X Terhadap Indonesia Emas 2045 Menggunakan Algoritma Support Vector Machine," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 7657–7665, 2024, doi: 10.36040/jati.v8i4.10327.

[5] 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.

[6] J. Y. Siahaan and A. Saifudin, "Penerapan Algoritma Naïve Bayes dalam Menganalisis Sentimen pada Review Pengguna E-Commerce," *Media Online)*, vol. 4, no. 1, pp. 305–316, 2023.

[7] I. B. Setiawan, J. Maulidar, and Nurchim, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Pada Aplikasi Kesehatan Digital," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 186–195, 2024, [Online]. Available: https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/1229

[8] R. Buyya, R. N. Calheiros, and A. V. Dastjerdi, *Big Data: Principle & Paradigms*, vol. 1, no. 69. 2016.

[9] M. Z. Haq, C. S. Octiva, A. Ayuliana, U. W. Nuryanto, and D. Suryadi, "Algoritma Naïve Bayes untuk Mengidentifikasi Hoaks di Media Sosial," *J. Minfo Polgan*, vol. 13, no. 1, pp. 1079–1084, 2024, doi: 10.33395/jmp.v13i1.13937.

[10] D. A. Punkastyo, F. Septian, and A. Syaripudin, "Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Siswa," *J. Syst. Comput. Eng.*, vol. 5, no. 1, pp. 24–35, 2024, doi: 10.61628/jsce.v5i1.1073.

[11] Asro, Sudaryono, and A. Sulaiman, "Analisis Sentimen tentang Transformasi Program Makan Siang menjadi Makan Bergizi Gratis menggunakan Logistik Regression pada laman Youtube," *J. ICT Inf. Commun. Technol.*, vol. 24, no. 1, 2024, [Online]. Available: https://ejournal.ikmi.ac.id/index.php/jict-ikmi

[12] S. S. Fasha and D. Tesniyadi, "Analisis Wacana Kritis Pada Artikel Tempo.co yang Berjudul "Dana BOS untuk Program Makan Siang Gratis"," vol. 4, pp. 15077–15089, 2024.

[13] U. Karomah, F. C. Wahyuni, and Y. D. Trisnasari, "Program Penyelenggaraan Makan Siang Sekolah: Studi Literatur tentang Dampak Kesehatan, Hambatan dan Tantangan," *Salus Cult. J. Pembang. Mns. dan Kebud.*, vol. 4, no. 1, 2024, doi: 10.55480/saluscultura.v4i1.188.

[14] A. A. Merlinda and Y. Yusuf, "Analisis Program Makan Gratis Prabowo Subianto Terhadap Strategi Peningkatan Motivasi Belajar Siswa di Sekolah Tinjauan dari Perspektif Sosiologi Pendidikan," vol. 7, no. 2, pp. 1364–1373, 2025.

[15] A. N. Ikhsan, P. Subarkah, Alifah Dafa Iftinani, and A. N. Fadilah, "Analisis Sentimen Kenaikan PPN menggunakan Algoritma Naïve Bayes dan Support Vector Machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.