



Contents lists available at [Journal IICET](#)

JRTI (Jurnal Riset Tindakan Indonesia)

ISSN: 2502-079X (Print) ISSN: 2503-1619 (Electronic)

Journal homepage: <https://jurnal.iicet.org/index.php/jrti>



Analisis Sentimen Wacana Publik di Media Sosial X terhadap Pengungsi Rohingya: Pendekatan Klasifikasi Naïve Bayes

Muhammad Chairul Azmi^{*)}, Muhammad Ikhsan
Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Article Info

Article history:

Received Oct 12th, 2025
Revised Nov 20th, 2025
Accepted Dec 26th, 2025

Keyword:

Analisis Sentimen,
Twitter,
Naïve Bayes,
Pengungsi Rohingya,
Public Perception,
Refugee Discourse,
Social Media Analytics

ABSTRACT

Kedatangan pengungsi etnis Rohingya di Indonesia memicu beragam respons publik yang secara luas diekspresikan melalui media sosial X. Dinamika opini tersebut mencerminkan interaksi kompleks antara nilai kemanusiaan, solidaritas sosial, serta kekhawatiran sosial dan ekonomi yang berkembang di masyarakat. Penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan sentimen opini masyarakat terhadap kedatangan pengungsi Rohingya di Indonesia dengan menggunakan algoritma Naïve Bayes berbasis data media sosial X. Hasil analisis menunjukkan bahwa mayoritas opini masyarakat cenderung bernada positif, yang merefleksikan adanya empati dan dukungan terhadap isu kemanusiaan pengungsi Rohingya. Namun demikian, proporsi sentimen negatif yang cukup signifikan juga ditemukan, terutama dipengaruhi oleh faktor sosial, pemberitaan media, serta penyebaran informasi yang tidak valid di ruang digital. Model klasifikasi yang dibangun mampu mencapai tingkat akurasi sebesar 80,13%, menunjukkan bahwa algoritma Naïve Bayes memiliki kinerja yang cukup baik dalam mengklasifikasikan sentimen teks berbahasa Indonesia. Temuan ini mengindikasikan bahwa meskipun narasi kemanusiaan masih mendominasi wacana publik, terdapat potensi pergeseran opini akibat isu sosial dan disinformasi yang berkembang di media sosial. Oleh karena itu, hasil penelitian ini dapat menjadi rujukan bagi pembuat kebijakan dalam merumuskan strategi komunikasi publik yang lebih efektif serta memperkuat pengelolaan isu pengungsi berbasis data dan persepsi masyarakat.



© 2025 The Authors. Published by IICET.

This is an open access article under the CC BY-NC-SA license
(<https://creativecommons.org/licenses/by-nc-sa/4.0>)

Corresponding Author:

Muhammad Chairul Azmi,
Universitas Islam Negeri Sumatera Utara
Email: muharriz1234@gmail.com

Introduction

Krisis kemanusiaan yang menimpa etnis Rohingya di Rakhine State, Myanmar, telah menjadi isu internasional yang memicu respons luas dari berbagai negara, termasuk Indonesia. Sejak meningkatnya eskalasi konflik pasca serangan Arakan Rohingya Salvation Army (ARSA) pada Agustus 2017, ratusan ribu pengungsi Rohingya terpaksa meninggalkan wilayah asal mereka dan mencari perlindungan ke negara-negara Asia Tenggara. Indonesia, meskipun bukan negara tujuan akhir dan tidak meratifikasi Konvensi Pengungsi 1951, secara konsisten menunjukkan sikap terbuka dengan memberikan penampungan sementara bagi pengungsi Rohingya, terutama di wilayah Aceh. Sikap ini mendapat apresiasi dari UNHCR dan mencerminkan komitmen Indonesia terhadap nilai-nilai kemanusiaan dan solidaritas regional (Moy & Kusuma, 2022).

Namun, seiring dengan berlanjutnya kedatangan pengungsi Rohingya sejak 2015 hingga beberapa tahun terakhir, dinamika respons masyarakat Indonesia menunjukkan perubahan yang signifikan. Pada fase awal,

kedatangan pengungsi umumnya disambut dengan empati dan solidaritas. Akan tetapi, dalam perkembangannya, muncul berbagai ketegangan sosial yang dipicu oleh keterbatasan fasilitas, perbedaan norma sosial, serta sejumlah insiden yang melibatkan oknum pengungsi. Peristiwa penolakan pengungsi Rohingya oleh sebagian masyarakat Aceh pada November 2023 menjadi indikator kuat adanya pergeseran sikap publik, dari solidaritas kemanusiaan menuju resistensi sosial (Indradipradana, 2023). Situasi ini semakin diperburuk oleh penyebaran hoaks, narasi kebencian, dan informasi yang tidak terverifikasi di media sosial, yang berpotensi membentuk persepsi publik secara negatif (Sopamena & Abbas, 2023).

Dalam konteks tersebut, media sosial khususnya Twitter menjadi ruang diskursif utama tempat masyarakat mengekspresikan opini, emosi, dan sikap mereka terhadap isu pengungsi Rohingya. Twitter tidak hanya berfungsi sebagai sarana komunikasi, tetapi juga sebagai arena pembentukan wacana publik (public discourse) yang dinamis dan responsif terhadap peristiwa-peristiwa tertentu. Oleh karena itu, analisis terhadap konten media sosial menjadi penting untuk memahami bagaimana persepsi publik terbentuk, bagaimana narasi positif dan negatif berkembang, serta bagaimana sentimen masyarakat bereaksi terhadap peristiwa sosial tertentu yang berkaitan dengan krisis pengungsi.

Meskipun sejumlah penelitian sebelumnya telah menerapkan analisis sentimen berbasis media sosial menggunakan berbagai algoritma klasifikasi, sebagian besar studi tersebut masih berfokus pada isu-isu umum seperti ulasan aplikasi, film, atau produk komersial. Penelitian-penelitian tersebut cenderung menekankan aspek performa algoritma, seperti akurasi, precision, dan recall, tanpa mengaitkannya secara mendalam dengan konteks sosial dan implikasi substantif dari sentimen yang dianalisis. Dalam konteks isu kemanusiaan dan pengungsi, kajian yang menghubungkan pendekatan komputasional dengan pertanyaan penelitian sosial—seperti bagaimana narasi dominan terbentuk, bagaimana sentimen publik berubah seiring waktu, dan bagaimana media sosial memengaruhi persepsi terhadap kelompok rentan masih relatif terbatas.

Dengan demikian, terdapat celah penelitian (research gap) yang perlu diisi, yaitu kurangnya studi yang secara sistematis menganalisis sentimen publik terhadap pengungsi Rohingya di Indonesia dengan menempatkan analisis komputasional sebagai alat untuk menjawab persoalan sosial yang lebih luas. Penelitian ini tidak hanya berupaya mengklasifikasikan sentimen positif dan negatif, tetapi juga memahami kecenderungan persepsi publik yang berkembang di ruang digital sebagai refleksi dari dinamika sosial yang terjadi di masyarakat.

Pemilihan algoritma Naïve Bayes dalam penelitian ini didasarkan pada pertimbangan metodologis dan empiris. Meskipun tergolong algoritma klasik, Naïve Bayes memiliki keunggulan dalam menangani data teks berdimensi tinggi, efisiensi komputasi, serta stabilitas kinerja pada dataset berukuran terbatas. Selain itu, sejumlah penelitian komparatif menunjukkan bahwa Naïve Bayes mampu menghasilkan performa yang kompetitif, bahkan dalam beberapa kasus melampaui metode yang lebih kompleks seperti Support Vector Machine, khususnya pada teks berbahasa Indonesia. Dengan karakteristik tersebut, Naïve Bayes dinilai relevan sebagai pendekatan awal (baseline model) yang kuat untuk menganalisis wacana publik di media sosial secara sistematis dan terukur.

Berdasarkan latar belakang tersebut, penelitian ini diarahkan untuk menjawab pertanyaan penelitian utama: bagaimana kecenderungan sentimen dan persepsi publik masyarakat Indonesia terhadap kedatangan pengungsi Rohingya yang tercermin dalam wacana di media sosial Twitter? Untuk menjawab pertanyaan tersebut, penelitian ini menerapkan analisis sentimen berbasis algoritma Naïve Bayes terhadap data tweet berbahasa Indonesia. Hasil penelitian diharapkan tidak hanya memberikan gambaran kuantitatif mengenai distribusi sentimen, tetapi juga menjadi dasar pemahaman yang lebih luas mengenai dinamika opini publik, sehingga dapat dimanfaatkan sebagai bahan pertimbangan dalam perumusan kebijakan dan strategi komunikasi publik terkait isu pengungsi dan krisis kemanusiaan.

Method

Penelitian ini dilakukan di Universitas Islam Negeri Sumatera Utara yang berada di Jl. Lap. Golf, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara 20353. Bagian metodologi dalam penelitian ini dirancang dengan perhatian khusus pada validitas konseptual proses analisis sentimen. Salah satu tahapan krusial adalah proses pelabelan data, yang tidak dilakukan sepenuhnya menggunakan pendekatan berbasis leksikon sebagai ground truth. Pendekatan leksikon dalam penelitian ini hanya dimanfaatkan sebagai alat bantu awal untuk memberikan gambaran awal kecenderungan sentimen, bukan sebagai sumber kebenaran utama. Pelabelan sentimen dilakukan melalui human annotation dengan melibatkan dua anotator independen yang memiliki pemahaman terhadap konteks sosial dan isu kemanusiaan pengungsi Rohingya. Setiap tweet diklasifikasikan ke dalam kategori positif, netral, dan negatif berdasarkan makna kontekstual dan sikap yang diekspresikan. Untuk menjamin reliabilitas pelabelan, tingkat kesepakatan antar anotator diuji menggunakan

Cohen's Kappa, sehingga label yang digunakan dalam proses pelatihan dan pengujian model merepresentasikan penilaian manusia yang independen dan dapat dipertanggungjawabkan secara metodologis.

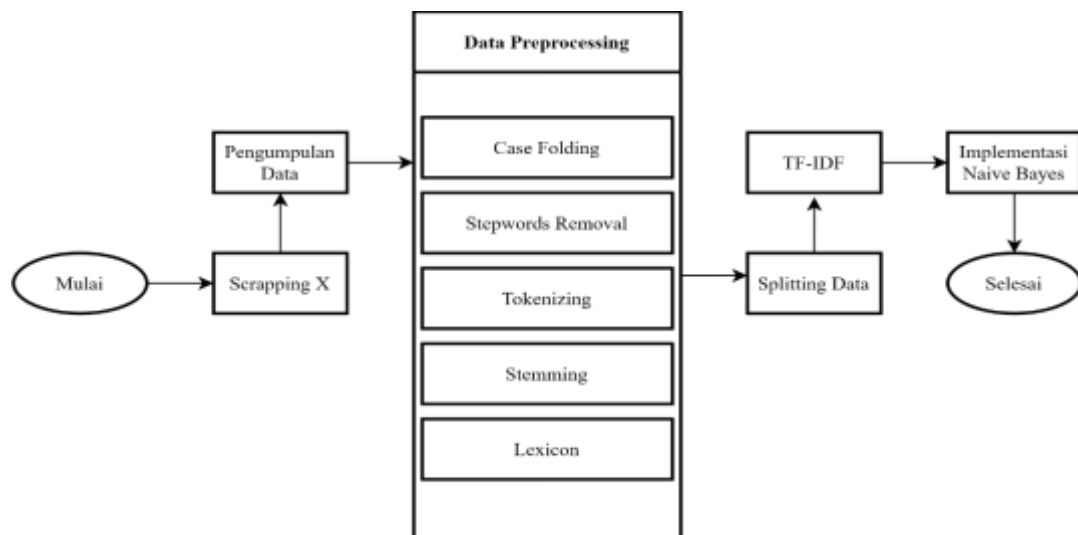
Pengumpulan data dilakukan secara transparan dan konsisten pada satu rentang waktu yang jelas, yaitu dari 1 Januari 2023 hingga 1 November 2024, dengan menggunakan alat Tweet-Harvest dan kata kunci yang relevan dengan isu Rohingya. Penegasan rentang waktu ini penting untuk memastikan bahwa data yang dianalisis merefleksikan dinamika wacana publik pada periode tertentu dan menghindari ambiguitas temporal dalam interpretasi hasil. Data yang diperoleh berupa tweet berbahasa Indonesia yang telah melalui proses seleksi untuk memastikan relevansi topik dan menghilangkan duplikasi.

Tahap pra-pemrosesan data tidak dilakukan sebagai prosedur teknis semata, melainkan dengan mempertimbangkan karakteristik kontekstual teks media sosial. Proses cleaning, case folding, tokenizing, stopword removal, stemming, dan normalisasi dilakukan secara selektif agar tidak menghilangkan elemen penting seperti kata bermuatan emosi, istilah populer, atau ekspresi informal yang berperan dalam pembentukan sentimen. Dengan pendekatan ini, makna sosial dan emosional yang terkandung dalam tweet tetap terjaga, sehingga hasil analisis sentimen tidak tereduksi menjadi sekadar representasi teknis yang kehilangan konteks.

Dengan dasar metodologis tersebut, data yang telah diberi label secara manual kemudian dibagi menjadi data pelatihan dan data pengujian dengan perbandingan 80:20, diekstraksi fiturnya menggunakan metode TF-IDF, dan diklasifikasikan menggunakan algoritma Naïve Bayes. Pendekatan ini memastikan bahwa model supervised learning diuji terhadap label yang bersumber dari penilaian manusia yang independen, bukan sekadar mereplikasi kesalahan dari sistem pelabelan otomatis, sehingga hasil klasifikasi yang diperoleh memiliki validitas empiris yang lebih kuat.

Untuk memperjelas alur dan tahapan penelitian yang dilakukan, diperlukan suatu kerangka penelitian yang menggambarkan proses penelitian secara keseluruhan. Kerangka penelitian ini disusun untuk menunjukkan tahapan-tahapan yang dilalui, mulai dari pengumpulan data, proses pengolahan dan analisis data, hingga penerapan algoritma Naïve Bayes dalam melakukan klasifikasi opini masyarakat.

Kerangka penelitian tersebut disajikan dalam bentuk diagram pada Gambar berikut, sehingga dapat memberikan gambaran yang lebih jelas dan sistematis mengenai alur penelitian yang dilakukan.



Gambar 1. Kerangka Penelitian

Berdasarkan kerangka penelitian pada Gambar 1, penelitian ini diawali dengan tahap pengumpulan data yang dilakukan melalui studi pustaka dan observasi terhadap opini masyarakat di media sosial X. Data dikumpulkan menggunakan alat Tweet-Harvest dengan kata kunci "Rohingya" pada periode 1 Januari 2023 hingga 1 November 2024, sehingga diperoleh 1.000 tweet berupa data tekstual yang relevan dengan topik penelitian. Selanjutnya, data yang terkumpul melalui tahap preprocessing untuk membersihkan dan menyiapkan teks, meliputi cleaning, case folding, stopword removal, tokenizing, stemming, serta normalisasi dan penggunaan leksikon sentimen. Data yang telah diproses kemudian dibagi menjadi data pelatihan dan data pengujian dengan perbandingan 80:20 untuk memastikan kinerja model dapat dievaluasi secara optimal. Tahap berikutnya adalah pembobotan kata menggunakan metode TF-IDF guna mengekstraksi fitur teks yang relevan. Seluruh fitur tersebut selanjutnya diklasifikasikan menggunakan algoritma Naïve Bayes untuk menghasilkan

klasifikasi sentimen opini masyarakat ke dalam kategori negatif, netral, dan positif, sekaligus mengukur tingkat akurasi model yang dihasilkan.

Results and Discussions

Crawling Data X

Dalam penelitian ini, proses penyiapan data berkaitan dengan data yang bersumber dari media sosial, yang merupakan salah satu sumber informasi paling dinamis terkait kedatangan etnis rohingya dan situasi yang tengah berlangsung di masyarakat. Namun, data dari media sosial tidak akan berguna jika belum melalui tahap pengolahan. Kegiatan pengolahan data, khususnya pengambilan data dari web atau media sosial untuk mengidentifikasi pola-pola tersembunyi di dalamnya, dikenal dengan istilah crawling (Yuniar et al., 2022).

Data dikumpulkan dari opini publik terkait kedatangan etnis rohingya dengan periode 2023–2024 yang disampaikan melalui media sosial X. Data yang digunakan berupa tweet berbahasa Indonesia. Pengambilan data dilakukan melalui metode crawling pada media sosial X dengan menggunakan library tweet-harvest. Proses ini dilakukan dengan menggunakan kata kunci "pengungsi Rohingya", "kapal Rohingya", "Rohingya Aceh", "pengungsi di Aceh".

```
# Crawling Data
filename = 'rohingnya.csv'
search_keyword = '"pengungsi Rohingya" OR "kapal Rohingya" OR "Rohingya Aceh" OR "pengungsi di Aceh"' since:2024-01-01 until:2025-12-31 lang:id'
limit = 1000
!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

Gambar 1. proses crawling

Kode yang ada pada gambar 1 digunakan untuk melakukan proses crawling, di mana tweet yang mengandung kata kunci seperti "pengungsi Rohingya", "kapal Rohingya", "Rohingya Aceh", dan "pengungsi di Aceh" dikumpulkan dalam rentang waktu 1 Januari 2024 hingga 31 Desember 2025. Data tersebut disimpan dalam format CSV dengan nama file rohingnya.csv dan akan digunakan sebagai dasar dalam analisis sentimen serta identifikasi isu-isu sosial yang berkembang di masyarakat terkait pengungsi Rohingya.

Data Preprocessing

Setelah data dikumpulkan, maka selanjutnya akan dilakukan pengolahan data teks, yang bertujuan untuk mengurangi gangguan atau noise dalam data sehingga dapat diproses lebih lanjut. Proses ini mencakup berbagai tahapan rutin yang diperlukan untuk mempersiapkan data sebelum digunakan. Dengan melalui proses ini, data teks akan dibersihkan sehingga menghasilkan dataset yang lebih terstruktur dan siap digunakan, yang pada akhirnya akan mempermudah tahapan pemrosesan selanjutnya (Putra et al., 2024). Kumpulan data rohingnya.csv sebelum dilakukan preprocessing pada tabel 1.

Tabel 1. Data rohingya.csv

Full_text	Created_at
Mereka membuat Hamas dan pejuang Palestina sebagai pelaku. Membuat pengungsi Rohingya tampak buruk sehingga tidak layak dibantu. Membuat Assad tampak menjadi pahlawan. Narasi2 seperti ini akan terus dipakai hingga dunia mendekati masa-nya.	"Sun Dec 29 09:11:11 +0000 2024"
@sjfi_twider Kalau menurut aku etnis Rohingya ini tetap akan Merdeka kerana Maungdaw belum sepenuhnya stabil sekarang dan beberapa tahun nanti. Betul krisis pengungsi Rohingya tetap lanjut cuma Rohingya ini akan jadi puak Rojava versi Asia.	"Sat Dec 28 02:57:10 +0000 2024"
@Y_D_Y_P Justru Rohingya ada di pihak junta kalau kelompok yang memberontak ini sangat rasis pada Rohingya Siap2 makin banyak pengungsi rohingya kalau gini	"Thu Dec 26 16:45:00 +0000 2024"
@AriSaputra81167 @idextratime Dress code nya pakaian para pengungsi Rohingya.	"Fri Apr 25 08:42:52 +0000 2025"
https://t.co/fLJqTRVyC2 Mantap ya pemerintah Indonesia memperlakukan pengungsi Rohingya diberikan lengkap fasilitas kemanusiaannya sementara untuk RAKYATNYA SENDIRI yang terkena musibah hanya diberi makan SEGELAS BUBUR KACANG IJO BAROKAH PRESIDEN @prabowo @gibran_tweet	"Fri Nov 29 19:53:43 +0000 2024"

Berikutnya dilakukan data preprocessing pada data rohingya.csv, tahapan dalam data preprocessing meliputi:

Cleaning Text

Cleaning adalah proses pembersihan teks dari karakter atau simbol yang tidak diperlukan. Langkah ini mencakup penghapusan tanda baca, angka, URL, karakter khusus, dan elemen-elemen lain yang tidak relevan (Kartika Sari et al., 2024). Berikut adalah hasil dari cleaning text:

Tabel 2. Contoh hasil cleaning text

Full_text	Preprocessing_text
Mereka membuat Hamas dan pejuang Palestina sebagai pelaku. Membuat pengungsi Rohingya tampak buruk sehingga tidak layak dibantu. Membuat Assad tampak menjadi pahlawan. Narasi2 seperti ini akan terus dipakai hingga dunia mendekati masa-nya.	Mereka membuat Hamas dan pejuang Palestina sebagai pelaku Membuat pengungsi Rohingya tampak buruk sehingga tidak layak dibantu Membuat Asad tampak menjadi pahlawan seperti ini akan terus dipakai hinga dunia mendekati
@sjfi_twider Kalau menurut aku etnis Rohingya ini tetap akan Merdeka kerana Maungdaw belum sepenuhnya stabil sekarang dan beberapa tahun nanti. Betul krisis pengungsi Rohingya tetap lanjut cuma Rohingya ini akan jadi puak Rojava versi Asia.	Kalau menurut aku etnis Rohingya ini tetap akan Merdeka kerana Maungdaw belum sepenuhnya stabil sekarang dan beberapa tahun nanti Betul krisis pengungsi Rohingya tetap lanjut cuma Rohingya ini akan jadi puak Rojava versi Asia
@Y_D_Y_P Justru Rohingya ada di pihak junta kalau kelompok yang memberontak ini sangat rasis pada Rohingya Siap2 makin banyak pengungsi rohingya kalau gini	Justru Rohingya ada di pihak junta kalau kelompok yang memberontak ini sangat rasis pada Rohingya makin banyak pengungsi rohingya kalau gini
@AriSaputra81167 @idextratime Dress code nya pakaian para pengungsi Rohingya.	Dres code nya pakaian para pengungsi Rohingya
https://t.co/fLJqTRVyC2 Mantap ya pemerintah Indonesia memperlakukan pengungsi Rohingya diberikan lengkap fasilitas kemanusiaannya sementara untuk RAKYATNYA SENDIRI yang terkena musibah hanya diberi makan SEGELAS BUBUR KACANG IJO BAROKAH PRESIDEN @prabowo @gibran_tweet	Mantap ya pemerintah Indonesia memperlakukan pengungsi Rohingya diberikan lengkap fasilitas sementara untuk RAKYATNYA SENDIRI yang terkena musibah hanya diberi makan SEGELAS BUBUR KACANG IJO BAROKAH PRESIDEN

Case Folding

Case folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil (lowercase) (Multi Fani & Santoso, 2023). Hal ini dilakukan untuk memastikan bahwa perbedaan dalam kapitalisasi tidak mempengaruhi analisis teks. Misalnya, "Data" dan "data" akan dianggap sama setelah case folding. Hal ini membantu dalam menyederhanakan data dan mengurangi variasi kata yang tidak diperlukan. Berikut adalah contoh hasil case folding text:

Tabel 3. Hasil case folding text

Full_text	Preprocessing_text
Mereka membuat Hamas dan pejuang Palestina sebagai pelaku. Membuat pengungsi Rohingya tampak buruk sehingga tidak layak dibantu. Membuat Assad tampak menjadi pahlawan. Narasi2 seperti ini akan terus dipakai hingga dunia mendekati masa-nya.	mereka membuat hamas dan pejuang palestina sebagai pelaku membuat pengungsi rohingya tampak buruk sehinga tidak layak dibantu membuat asad tampak menjadi pahlawan seperti ini akan terus dipakai hinga dunia mendekati
@sjfi_twider Kalau menurut aku etnis Rohingya ini tetap akan Merdeka kerana Maungdaw belum sepenuhnya stabil sekarang dan beberapa tahun nanti. Betul krisis pengungsi Rohingya tetap lanjut cuma Rohingya ini akan jadi puak Rojava versi Asia.	kalau menurut aku etnis rohingya ini tetap akan merdeka kerana maungdaw belum sepenuhnya stabil sekarang dan beberapa tahun nanti betul krisis pengungsi rohingya tetap lanjut cuma rohingya ini akan jadi puak rojava versi asia
@Y_D_Y_P Justru Rohingya ada di pihak junta kalau kelompok yang memberontak ini sangat rasis pada Rohingya Siap2 makin banyak pengungsi rohingya kalau gini	justru rohingya ada di pihak junta kalau kelompok yang memberontak ini sangat

Full_text	Preprocessing_text
@AriSaputra81167 @idextratime Dress code nya pakaian para pengungsi Rohingya. https://t.co/fLJqTRVyC2 Mantap ya pemerintah Indonesia memperlakukan pengungsi Rohingya diberikan lengkap fasilitas kemanusiaannya sementara untuk RAKYATNYA SENDIRI yang terkena musibah hanya diberi makan SEGELAS BUBUR KACANG IJO BAROKAH PRESIDEN @prabowo @gibrantweet	rasis pada rohingya makin banyak pengungsi rohingya kalau gini dres code nya pakaian para pengungsi rohingya mantap ya pemerintah indonesia memperlakukan pengungsi rohingya diberikan lengkap fasilitas sementara untuk rakyatnya sendiri yang terkena musibah hanya diberi makan segelas bubur kacang ijo barokah presiden

Stepwords Removal

Stopword removal merupakan tahap dalam pemrosesan teks yang bertujuan untuk menghilangkan kata-kata umum yang sering muncul namun tidak memiliki makna signifikan. Contohnya meliputi kata-kata seperti "dan", "atau", "yang", "itu" dalam bahasa Indonesia, atau "the", "is", "in", "and" dalam bahasa Inggris. Proses ini berguna untuk mengurangi fitur yang kurang relevan sehingga analisis dapat lebih berfokus pada kata-kata yang mengandung informasi penting (Shah et al., 2024).

Tabel 4. Hasil case folding

Full_text	Preprocessing_text
Mereka membuat Hamas dan pejuang Palestina sebagai pelaku. Membuat pengungsi Rohingya tampak buruk sehingga tidak layak dibantu. Membuat Assad tampak menjadi pahlawan. Narasi2 seperti ini akan terus dipakai hingga dunia mendekati masa-nya.	[hamas,pejuang,palestina,pelaku,pengungsi,rohingya,buruk,sehinga,layak,dibantu,asad,pahlawan,dipakai,hinga,dunia,mendekati]
@sjfi_twider Kalau menurut aku etnis Rohingya ini tetap akan Merdeka kerana Maungdaw belum sepenuhnya stabil sekarang dan beberapa tahun nanti. Betul krisis pengungsi Rohingya tetap lanjut cuma Rohingya ini akan jadi puak Rojava versi Asia.	[etnis,rohingya,merdeka,kerana,maungdaw,sepenuhnya,stabil,krisis,pengungsi,rohingya,rohingya,puak,rojava,versi,asia]
@Y_D_Y_P Justru Rohingya ada di pihak junta kalau kelompok yang memberontak ini sangat rasis pada Rohingya Siap2 makin banyak pengungsi rohingya kalau gini	[rohingya,junta,kelompok,memberontak,rasis,rohingya,pengungsi,rohingya,gini]
@AriSaputra81167 @idextratime Dress code nya pakaian para pengungsi Rohingya. https://t.co/fLJqTRVyC2 Mantap ya pemerintah Indonesia memperlakukan pengungsi Rohingya diberikan lengkap fasilitas kemanusiaannya sementara untuk RAKYATNYA SENDIRI yang terkena musibah hanya diberi makan SEGELAS BUBUR KACANG IJO BAROKAH PRESIDEN @prabowo @gibrantweet	[dres,code,nya,pakaian,pengungsi,rohingya] [mantap,ya,pemerintah,indonesia,memperlakukan,pengungsi,rohingya,lengkap,fasilitas,rakyatnya,terkena,musibah,makan,segelas,bubur,kacang,ijo,barokah,presiden]

Token

Tokenizing merupakan proses yang bertujuan untuk memecah teks menjadi bagian-bagian kecil, seperti kata atau frasa, dalam sebuah kalimat. Berikut ini adalah contoh dari proses tokenizing tabel 5.

Tabel 5. Hasil tokenizing

Full_text	Preprocessing_text
Mereka membuat Hamas dan pejuang Palestina sebagai pelaku. Membuat pengungsi Rohingya tampak buruk sehingga tidak layak dibantu. Membuat Assad tampak menjadi pahlawan. Narasi2 seperti ini akan terus dipakai hingga dunia mendekati masa-nya.	['hamas', 'juang', 'palestina', 'laku', 'ungsi', 'rohingya', 'buruk', 'sehinga', 'layak', 'bantu', 'asad', 'pahlawan', 'pakai', 'hinga', 'dunia', 'dekat']
@sjfi_twider Kalau menurut aku etnis Rohingya ini tetap akan Merdeka kerana Maungdaw belum sepenuhnya stabil sekarang dan beberapa tahun nanti. Betul krisis pengungsi Rohingya tetap lanjut cuma Rohingya ini akan jadi puak Rojava versi Asia.	['etnis', 'rohingya', 'merdeka', 'rana', 'maungdaw', 'sepenuh', 'stabil', 'krisis', 'ungsi', 'rohingya', 'rohingya', 'puak', 'rojava', 'versi', 'asia']
@Y_D_Y_P Justru Rohingya ada di pihak junta kalau kelompok yang memberontak ini sangat rasis pada Rohingya Siap2 makin banyak pengungsi rohingya kalau gini	['rohingya', 'junta', 'kelompok', 'berontak', 'ras', 'rohingya', 'ungsi', 'rohingya', 'gin']
@AriSaputra81167 @idextratime Dress code nya pakaian para pengungsi Rohingya.	['dres', 'code', 'nya', 'pakai', 'ungsi', 'rohingya']
Mantap ya pemerintah Indonesia memperlakukan pengungsi Rohingya diberikan lengkap fasilitas kemanusiaannya sementara untuk RAKYATNYA SENDIRI yang terkena PRESIDEN @prabowo @gibran_tweet	['mantap', 'ya', 'perintah', 'indonesia', 'laku', 'ungsi', 'rohingya', 'lengkap', 'fasilitas', 'rakyat', 'kena', 'musibah', 'makan', 'gelas', 'bubur', 'kacang', 'ijo', 'barokah', 'presiden']

Stemming

Stemming merupakan proses mengubah kata menjadi bentuk dasarnya (stem) dengan cara menghapus awalan, akhiran, atau imbuhan lainnya. Sebagai contoh, dalam bahasa Inggris, kata-kata seperti "running", "runner", dan "ran" akan diubah menjadi "run". Sementara dalam bahasa Indonesia, kata seperti "bermain", "bermainan", dan "dimainkan" akan direduksi menjadi "main". Proses stemming membantu mengurangi ragam bentuk kata serta menyatukan kata-kata yang memiliki arti dasar yang serupa (Giovani et al., 2020). Berikut adalah contoh hasil dari stemming:

Tabel 6. Hasil stemming

Full_text	Preprocessing_text
Mereka membuat Hamas dan pejuang Palestina sebagai pelaku. Membuat pengungsi Rohingya tampak buruk sehingga tidak layak dibantu. Membuat Assad tampak menjadi pahlawan. Narasi2 seperti ini akan terus dipakai hingga dunia mendekati masa-nya.	hamas juang palestina laku ungsi rohingya buruk sehinga layak bantu asad pahlawan pakai hinga dunia dekat
@sjfi_twider Kalau menurut aku etnis Rohingya ini tetap akan Merdeka kerana Maungdaw belum sepenuhnya stabil sekarang dan beberapa tahun nanti. Betul krisis pengungsi Rohingya tetap lanjut cuma Rohingya ini akan jadi puak Rojava versi Asia.	etnis rohingya merdeka rana maungdaw sepenuh stabil krisis ungsi rohingya rohingya puak rojava versi asia
@Y_D_Y_P Justru Rohingya ada di pihak junta kalau kelompok yang memberontak ini sangat rasis pada Rohingya Siap2 makin banyak pengungsi rohingya kalau gini	rohingya junta kelompok berontak ras rohingya ungsi rohingya gin
Full text	Preprocessing_text
@AriSaputra81167 @idextratime Dress code nya pakaian para pengungsi Rohingya.	dres code nya pakai ungsi rohingya
https://t.co/fLJqTRVyC2 Mantap ya pemerintah Indonesia memperlakukan pengungsi Rohingya diberikan lengkap fasilitas kemanusiaannya sementara untuk RAKYATNYA SENDIRI yang terkena musibah hanya diberi makan SEGELAS BUBUR KACANG IJO BAROKAH PRESIDEN @prabowo @gibran_tweet	mantap ya perintah indonesia laku ungsi rohingya lengkap fasilitas rakyat kena musibah makan gelas bubur kacang ijo barokah presiden

Lexicon

Lexicon digunakan dalam proses pelabelan data setelah melalui tahapan pembersihan (preprocessing) teks (Ismail & Raden Bagus Fajriya Hakim, 2023). Pada tahap ini, setiap kalimat dianalisis berdasarkan keberadaan kata-kata yang terdapat dalam kamus lexicon, yang berisi daftar kata-kata bermuatan positif dan negatif. Setiap kata yang dikenali akan diberi skor sesuai dengan nilai sentimennya, dan skor total dari suatu kalimat dihitung berdasarkan jumlah serta jenis kata yang ditemukan.

Dengan menggunakan pendekatan ini, sistem dapat mengklasifikasikan teks ke dalam kategori sentimen seperti positif, negatif, atau netral tanpa perlu pelatihan model terlebih dahulu, karena penilaian didasarkan pada bobot kata yang telah ditentukan dalam kamus lexicon (Syakur, 2021). Pendekatan ini banyak digunakan dalam analisis sentimen yang berbasis aturan (rule-based). Berikut adalah contoh hasil dari pembobotan skor sentimen dengan lexicon:

Tabel 7. Hasil pembobotan dengan lexicon

Preprocessing_text	Word_sentiment	Sentiment_score	Sentiment_labe
hamas juang palestina laku ungsi rohingya buruk sehinga layak bantu asad pahlawan pakai hinga dunia dekat	{'hamas': 0, 'juang': 0, 'palestina': 0, 'laku': 0, 'ungsi': 0, 'rohingya': 0, 'buruk': -1, 'sehinga': 0, 'layak': 0, 'bantu': 0, 'asad': 0, 'pahlawan': 0, 'pakai': 0, 'hinga': 0, 'dunia': 0, 'dekat': 0}	-1	negative
etnis rohingya merdeka rana maungdaw sepenuh stabil krisis ungsi rohingya rohingya puak rojava versi asia	{'etnis': 0, 'rohingya': 0, 'merdeka': 0, 'rana': 0, 'maungdaw': 0, 'sepenuh': 0, 'stabil': 0, 'krisis': -1, 'ungsi': 0, 'puak': 0, 'rojava': 0, 'versi': 0, 'asia': 0}	-1	negative
rohingya junta kelompok berontak ras rohingya ungsi rohingya gin	{'rohingya': 0, 'junta': 0, 'kelompok': 0, 'berontak': 0, 'ras': 0, 'ungsi': 0, 'gin': 0}	0	Netral
dres code nya pakai ungsi Rohingya	{'dres': 0, 'code': 0, 'nya': 0, 'pakai': 0, 'ungsi': 0, 'rohingya': 0}	0	Netral
Preprocessing_text	Word_sentiment	Sentiment_score	Sentiment_labe
mantap ya perintah indonesia laku ungsi rohingya lengkap fasilitas rakyat kena musibah makan gelas bubur kacang ijo barokah presiden	{'mantap': 1, 'ya': 0, 'perintah': 0, 'indonesia': 0, 'laku': 0, 'ungsi': 0, 'rohingya': 0, 'lengkap': 0, 'fasilitas': 0, 'rakyat': 0, 'kena': 0, 'musibah': 0, 'makan': 0, 'gelas': 0, 'bubur': 0, 'kacang': 0, 'ijo': 0, 'barokah': 0, 'presiden': 0}	1	positive

TF-IDF

Pembobotan merupakan proses pemberian nilai pada suatu kata (term) dalam sebuah dokumen berdasarkan seberapa sering kata tersebut muncul. Nilai bobot ini menunjukkan seberapa penting suatu kata dalam konteks dokumen tersebut, di mana kata yang lebih sering muncul akan mendapatkan bobot yang lebih tinggi. Pada tahap ini, setiap tweet yang telah melalui proses preprocessing dan pelabelan dengan pendekatan lexicon akan diberikan bobot term. Tujuan dari pembobotan ini adalah agar metode Term Frequency-Inverse Document Frequency (TF-IDF) dapat digunakan untuk mengukur dan mengekstraksi relevansi kata-kata dalam dokumen, sehingga menghasilkan perhitungan bobot yang mencerminkan pentingnya setiap kata dalam keseluruhan isi dokumen.

Term Frequency (TF)

Pertama, kita beri label untuk tiap kalimat agar mudah dibaca.

Tabel 8. Label kalimat TF

Dokumen	Kalimat
D1	hamas juang palestina laku ungsi rohingya buruk seHINGA layak bantu asad pahlawan pakai hinga dunia dekat
D2	etnis rohingya merdeka rana maungdaw sepenuh stabil krisis ungsi rohingya rohingya puak rojava versi asia
D3	rohingya junta kelompok berontak ras rohingya ungsi rohingya gin
D4	dres code nya pakai ungsi rohingya mantap ya pemerintah indonesia laku ungsi rohingya
D5	lengkap fasilitas rakyat kena musibah makan gelas bubur kacang ijo barokah presiden

Kemudian kita menggunakan persamaan:

$$TF(t, d) = \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Jumlah total kata dalam dokumen } d}$$

Langkah selanjutnya adalah menghitung nilai TF untuk kata "rohingya":

Tabel 9. Nilai TF "rohingya"

Dokumen	TF
D1	1/5 = 0.0667
D2	3/17 = 0.176
D3	3/8 = 0.375
D4	1/6 = 0.1667
D5	1/17 = 0.0588

Tabel di atas menunjukkan nilai **Term Frequency (TF)** untuk kata "rohingya" yang terdapat dalam lima dokumen (D1 hingga D5). Nilai TF dihitung dengan membagi jumlah kemunculan kata "rohingya" dalam satu dokumen dengan total jumlah kata dalam dokumen tersebut. Sebagai contoh, pada Dokumen 1 (D1), kata "rohingya" muncul satu kali dari total lima kata penting yang dipertimbangkan, sehingga menghasilkan TF sebesar 0.0667. Nilai TF tertinggi terdapat pada Dokumen 3 (D3), yaitu sebesar 0.375, menunjukkan bahwa kata "rohingya" lebih dominan digunakan dalam dokumen tersebut dibandingkan dokumen lainnya.

Document Frequency (DF) dan Inverse Document Frequency (IDF)

$$IDF(t) = \log \frac{N}{1 + DF(t)}$$

Keterangan:

N = 5 (jumlah total dokumen)

DF(t) = banyaknya dokumen yang mengandung term t

Kata "Rohingya" muncul pada setiap dokumen, maka DF = 5:

$$IDF(rohingya) = \log \frac{5}{1 + 5} = \log \frac{5}{6} = -0.079$$

Maka didapatkan hasil IDF untuk kata "Rohingya" adalah -0.079.

TF-IDF

Nilai TF-IDF dapat dihitung dengan persamaan berikut:

$$TF - IDF = TF \times DF$$

Sehingga nilai TF-IDF untuk kata "Rohingya" dapat dilihat pada tabel berikut:

Tabel 10. Nilai TF-IDF kata “rohingya”

Dokumen	TF	IDF = -0.079	TF-IDF
D1	$1/15 = 0.0667$	-0.079	= -0.0053
D2	$3/17 = 0.176$	-0.079	= -0.0139
D3	$3/8 = 0.375$	-0.079	= -0.0296
D4	$1/6 = 0.1667$	-0.079	= -0.0132
D5	$1/17 = 0.0588$	-0.079	= -0.0046

Splitting Data

Proses splitting data bertujuan untuk membagi dataset menjadi dua bagian utama, yaitu data latih (training set) dan data uji (testing set).

```
X = df['processed_text']
y = df['sentiment']
vectorizer = CountVectorizer()
X_vect = vectorizer.fit_transform(X)
X_train, X_test, y_train, y_test = train_test_split(X_vect, y, test_size=0.2, random_state=42)
```

Gambar 2 , fungsi train_test_split

Dalam kode yang digunakan, fungsi train_test_split dari pustaka Scikit-learn digunakan untuk membagi data yang telah melalui proses vektorisasi (X_vect) dan label sentimen (y). Parameter test_size=0.2 menunjukkan bahwa 20% dari total data akan digunakan sebagai data uji, sementara 80% sisanya digunakan untuk melatih model. Pembagian ini penting agar model dapat dievaluasi menggunakan data yang belum pernah dilihat sebelumnya, sehingga dapat mengukur kemampuan generalisasi model terhadap data baru. Penggunaan random_state=42 memastikan bahwa pembagian data bersifat konsisten setiap kali kode dijalankan, sehingga hasil eksperimen dapat direproduksi dengan hasil yang sama.

Implementasi Naive Bayes

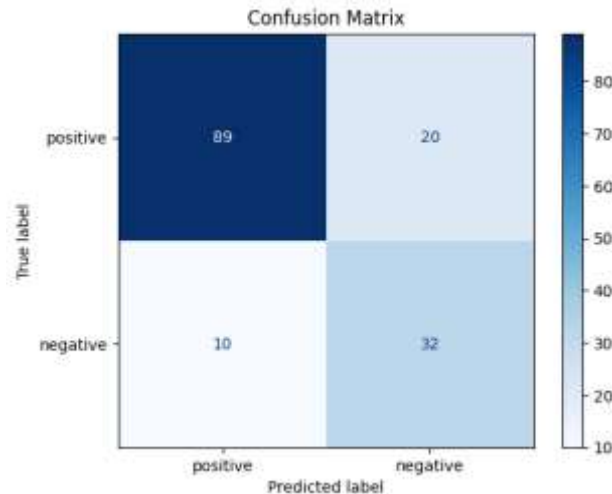
Klasifikasi Naive Bayes model Multinomial merupakan salah satu metode pembelajaran probabilistik didasarkan pada teorema Bayes. Algoritma ini bekerja pada konsep term frequency yang berarti berapa kali kata tersebut muncul dalam sebuah dokumen, sehingga setiap data perlu diberikan label sebelum dilakukan training.

```
model = MultinomialNB()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
# Confusion Matrix
cm = confusion_matrix(y_test, y_pred, labels=['positive', 'negative'])
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=['positive', 'negative'])
disp.plot(cmap='Blues')
plt.title("Confusion Matrix")
plt.show()

# Classification Report
print(classification_report(y_test, y_pred))
```

Gambar 3 penerapan klasifikasi model naive baye

Evaluasi model dilakukan dengan menggunakan data uji yang telah dipisahkan sebelumnya. Model Naive Bayes dilatih menggunakan fungsi fit() pada data latih, kemudian digunakan untuk memprediksi sentimen dari data uji dengan fungsi predict(). Hasil prediksi tersebut dibandingkan dengan label sebenarnya untuk menghitung confusion matrix, yang menggambarkan jumlah prediksi benar dan salah untuk masing-masing kelas. Visualisasi confusion matrix ditampilkan menggunakan ConfusionMatrixDisplay agar lebih mudah dianalisis. Selain itu, laporan klasifikasi ditampilkan melalui fungsi classification_report, yang menyajikan metrik evaluasi seperti precision, recall, F1-score, dan support untuk masing-masing kelas. Evaluasi ini bertujuan untuk menilai sejauh mana model mampu mengklasifikasikan sentimen secara akurat dan seimbang, terutama jika terdapat perbedaan distribusi antara kelas positif dan negatif.



Gambar 4. Confusion Matrix

Gambar di atas menunjukkan confusion matrix dari hasil evaluasi model Naive Bayes dalam melakukan klasifikasi sentimen. Dari total data uji, terdapat 89 data yang memiliki sentimen positif berhasil diklasifikasikan dengan benar sebagai positif (true positive), sedangkan 20 data positif diklasifikasikan secara keliru sebagai negatif (false negative). Untuk sentimen negatif, terdapat 32 data yang berhasil diprediksi dengan benar sebagai negatif (true negative), sementara 10 data negatif justru salah diklasifikasikan sebagai positif (false positive). Hasil ini menunjukkan bahwa model cenderung lebih akurat dalam mengklasifikasikan sentimen positif dibandingkan sentimen negatif, dengan jumlah kesalahan prediksi yang lebih rendah pada kelas negatif. Secara keseluruhan, visualisasi ini memberikan gambaran yang jelas mengenai performa dan distribusi kesalahan dari model klasifikasi yang digunakan.

	precision	recall	f1-score	support
negative	0.62	0.76	0.68	42
positive	0.90	0.82	0.86	109
accuracy			0.80	151
macro avg	0.76	0.79	0.77	151
weighted avg	0.82	0.80	0.81	151

Gambar 5. Evaluasi Model Naive Bayes

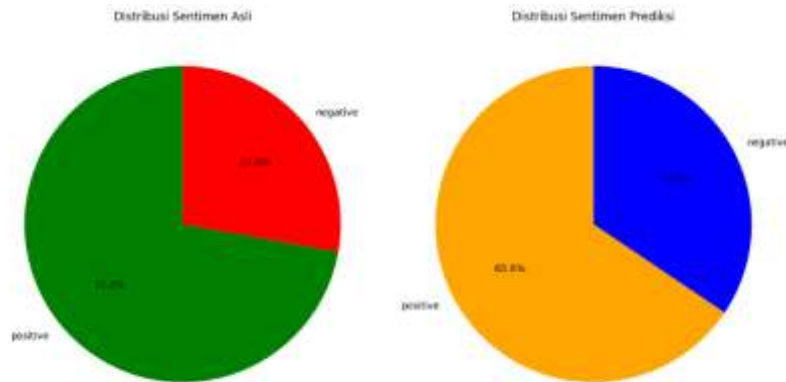
Gambar 3 merupakan hasil evaluasi performa model klasifikasi dalam bentuk metrik klasifikasi, yaitu precision, recall, f1-score, dan support untuk dua kelas: *negative* dan *positive*. Model menunjukkan performa yang lebih baik dalam mengklasifikasi kelas *positive* dengan precision sebesar 0.90, recall 0.82, dan f1-score 0.86 dari total 109 data. Sebaliknya, pada kelas *negative*, performa model menurun dengan precision hanya sebesar 0.62, recall 0.76, dan f1-score 0.68 dari 42 data. Secara keseluruhan, akurasi model mencapai 0.80. Rata-rata makro (macro avg), yang menghitung rata-rata metrik antar kelas secara seimbang, menghasilkan precision 0.76, recall 0.79, dan f1-score 0.77. Sementara itu, rata-rata berbobot (weighted avg), yang mempertimbangkan proporsi jumlah data tiap kelas, menunjukkan nilai precision 0.82, recall 0.80, dan f1-score 0.81. Hal ini menunjukkan bahwa meskipun model cukup baik secara keseluruhan, performanya kurang seimbang antara kedua kelas, dengan kecenderungan lebih baik dalam mengidentifikasi kelas *positive*.

```
[5] from sklearn.metrics import accuracy_score
# Akurasi
accuracy = accuracy_score(y_test, y_pred)
print(f"Akurasi Model: {accuracy:.2%}")

Akurasi Model: 80.13%
```

Gambar 6. Akurasi Model Naive Bayes

Nilai akurasi dihitung dan disimpan dalam variabel accuracy. Nilai tersebut kemudian dicetak ke layar dalam format persentase dengan dua angka di belakang koma menggunakan f-string. Output yang dihasilkan menunjukkan bahwa model memiliki tingkat akurasi sebesar 80.13%, yang berarti sekitar 80 dari 100 prediksi model sesuai dengan label yang benar. Ini mengindikasikan bahwa model memiliki performa klasifikasi yang cukup baik, meskipun evaluasi lebih mendalam dengan metrik lain tetap diperlukan untuk gambaran performa yang lebih menyeluruh.



Gambar 7. Perbandingan Actual vs Prediksi menggunakan Model Naive Bayes

Gambar tersebut menampilkan diagram perbandingan distribusi sentimen asli dan distribusi sentimen hasil prediksi model. Diagram di sebelah kiri menunjukkan distribusi sentimen asli, di mana sebanyak 72.2% data termasuk dalam kategori positive (ditampilkan dengan warna hijau), dan 27.8% termasuk kategori negative (ditampilkan dengan warna merah). Sementara itu, diagram di sebelah kanan menggambarkan distribusi sentimen hasil prediksi model. Di sini, persentase positive sedikit menurun menjadi 65.6% (berwarna oranye), sedangkan kategori negative meningkat menjadi 34.4% (berwarna biru). Perbedaan antara kedua distribusi ini mengindikasikan bahwa model sedikit cenderung memprediksi lebih banyak sentimen negatif dibandingkan proporsi sebenarnya, yang bisa menjadi sinyal bahwa model agak bias terhadap kelas negative, atau mengalami ketidakseimbangan dalam proses pelatihan atau prediksi.

Berdasarkan hasil analisis sentimen menggunakan model klasifikasi, ditemukan bahwa mayoritas respons masyarakat di media sosial terhadap kedatangan etnis Rohingya bernada positif. Hal ini terlihat dari distribusi sentimen asli, di mana sebanyak 72.2% data tergolong sebagai sentimen positive, sedangkan hanya 27.8% yang bersentimen negative. Ini mengindikasikan bahwa secara umum, masyarakat menunjukkan empati, dukungan, atau sikap terbuka terhadap kehadiran etnis Rohingya.

Hasil prediksi model juga mendukung temuan ini, meskipun terjadi sedikit pergeseran. Model memprediksi 65.6% data sebagai sentimen positive dan 34.4% sebagai negative. Perbedaan ini menunjukkan bahwa meskipun model cenderung sedikit melebihkan jumlah sentimen negatif, dominasi sentimen positif tetap konsisten. Lebih lanjut, model klasifikasi menunjukkan akurasi sebesar 80.13%, dengan performa lebih kuat dalam mengenali sentimen positive (precision 0.90, recall 0.82, f1-score 0.86) dibandingkan sentimen negative (precision 0.62, recall 0.76, f1-score 0.68). Hal ini menandakan bahwa model cenderung lebih akurat dalam mengidentifikasi dukungan atau simpati terhadap Rohingya dibandingkan kritik atau penolakan.

Dengan demikian, dapat disimpulkan bahwa secara umum, pandangan masyarakat di media sosial terhadap kedatangan etnis Rohingya cenderung positif, meskipun masih terdapat sebagian kelompok yang menyampaikan sentimen negatif. Sentimen negatif tersebut perlu dikaji lebih lanjut untuk memahami faktor-faktor yang melatarbelakanginya, baik dari sisi sosial, politik, maupun ekonomi.

Pembahasan

Hasil analisis sentimen terhadap 1.000 tweet berbahasa Indonesia yang membahas kedatangan pengungsi Rohingya menunjukkan bahwa wacana publik di media sosial Twitter didominasi oleh sentimen positif. Distribusi sentimen asli memperlihatkan bahwa sebesar 72,2% opini masyarakat bernada positif, sementara 27,8% bernada negatif. Temuan ini mengindikasikan bahwa, secara umum, masyarakat Indonesia masih menunjukkan empati dan sikap kemanusiaan terhadap pengungsi Rohingya, meskipun isu ini juga memunculkan resistensi dari sebagian kelompok.

Dominasi sentimen positif tersebut sejalan dengan karakter Indonesia sebagai negara yang menjunjung tinggi nilai kemanusiaan dan solidaritas sosial, sebagaimana tercermin dalam berbagai respons publik yang mendukung pemberian bantuan, perlindungan, dan fasilitas kemanusiaan bagi pengungsi. Kata-kata seperti “mantap”, “fasilitas”, dan “kemanusiaan” yang muncul dalam tweet bersentimen positif menunjukkan bahwa sebagian masyarakat memandang kebijakan pemerintah terhadap pengungsi Rohingya sebagai tindakan yang patut diapresiasi. Temuan ini memperkuat argumen bahwa media sosial tidak hanya menjadi ruang kritik, tetapi juga arena ekspresi dukungan moral terhadap isu-isu kemanusiaan.

Namun demikian, keberadaan sentimen negatif yang mencapai hampir sepertiga dari total data menunjukkan adanya polarisasi opini publik. Sentimen negatif umumnya berkaitan dengan narasi ketidakadilan sosial, kecemburuan ekonomi, serta persepsi bahwa pengungsi Rohingya mendapatkan perlakuan yang lebih baik dibandingkan masyarakat lokal yang mengalami kesulitan. Narasi semacam ini tampak jelas pada tweet yang membandingkan bantuan kepada pengungsi dengan kondisi warga terdampak bencana di dalam negeri. Temuan ini menunjukkan bahwa sentimen negatif tidak selalu berangkat dari penolakan kemanusiaan, melainkan lebih sering dipicu oleh ketidakpuasan terhadap kebijakan dan distribusi sumber daya.

Dari sisi performa model, algoritma Naïve Bayes menghasilkan akurasi sebesar 80,13%, yang menunjukkan bahwa model mampu mengklasifikasikan sentimen opini publik dengan tingkat ketepatan yang cukup baik. Model menunjukkan performa yang lebih kuat dalam mengidentifikasi sentimen positif, dengan nilai precision sebesar 0,90 dan f1-score 0,86, dibandingkan sentimen negatif yang memiliki precision 0,62 dan f1-score 0,68. Ketimpangan ini mengindikasikan bahwa ekspresi sentimen negatif di media sosial cenderung lebih beragam, implisit, dan kontekstual, sehingga lebih sulit dikenali oleh model berbasis frekuensi kata. Sebaliknya, sentimen positif umumnya diekspresikan secara lebih eksplisit melalui kata-kata bernada dukungan atau apresiasi, sehingga lebih mudah terdeteksi.

Perbedaan distribusi antara sentimen asli dan hasil prediksi model, di mana model cenderung sedikit melebihkan proporsi sentimen negatif, juga menunjukkan adanya bias klasifikasi yang umum terjadi pada dataset dengan distribusi kelas yang tidak seimbang. Meski demikian, dominasi sentimen positif tetap konsisten baik pada data aktual maupun hasil prediksi, sehingga kesimpulan substantif penelitian ini tidak berubah.

Jika dibandingkan dengan penelitian terdahulu, temuan ini memiliki beberapa kesesuaian sekaligus perbedaan yang menarik. Penelitian oleh Ratnawati (2020) yang menganalisis sentimen opini film di Twitter menggunakan algoritma Naïve Bayes menunjukkan bahwa model cenderung lebih akurat dalam mengklasifikasikan sentimen positif dibandingkan negatif, pola yang juga ditemukan dalam penelitian ini. Hal ini mengindikasikan bahwa karakteristik bahasa di Twitter yang sering kali singkat, informal, dan emosional memang lebih mendukung deteksi sentimen positif yang eksplisit.

Sementara itu, penelitian Nurzaman et al. (2022) mengenai analisis sentimen ulasan aplikasi menggunakan Naïve Bayes juga melaporkan akurasi yang berada pada kisaran 70–75%, sedikit lebih rendah dibandingkan akurasi yang diperoleh dalam penelitian ini. Perbedaan tersebut dapat disebabkan oleh konteks data yang berbeda, di mana isu kemanusiaan seperti pengungsi Rohingya cenderung memunculkan kata-kata emosional yang lebih kuat dan berulang, sehingga memudahkan proses klasifikasi berbasis teks.

Di sisi lain, penelitian-penelitian sebelumnya yang membahas pengungsi Rohingya lebih banyak dilakukan secara kualitatif atau deskriptif, dengan menekankan aspek kebijakan dan hubungan internasional. Penelitian ini melengkapi kajian tersebut dengan pendekatan analisis komputasional, yang memungkinkan pengukuran persepsi publik secara kuantitatif dan berbasis data besar. Dengan demikian, penelitian ini mengisi celah (research gap) antara kajian sosial kemanusiaan dan metode analitik berbasis data dalam memahami wacana publik mengenai krisis pengungsi.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa meskipun terdapat narasi negatif yang berkembang di media sosial, persepsi publik masyarakat Indonesia terhadap pengungsi Rohingya masih cenderung positif. Temuan ini penting bagi pembuat kebijakan dan pemangku kepentingan untuk memahami bahwa resistensi publik yang muncul tidak bersifat homogen, melainkan dipengaruhi oleh faktor sosial, ekonomi, dan informasi yang beredar di media. Oleh karena itu, strategi komunikasi publik yang lebih transparan dan edukatif diperlukan untuk meredam disinformasi serta menjaga dukungan masyarakat terhadap kebijakan penanganan pengungsi Rohingya.

Conclusions

Penelitian ini bertujuan untuk mengevaluasi bagaimana analisis sentimen berbasis komputasi dapat digunakan untuk memahami persepsi publik Indonesia terhadap isu sensitif, khususnya kedatangan pengungsi Rohingya, melalui wacana yang berkembang di media sosial Twitter. Berdasarkan hasil analisis, penelitian ini menyimpulkan dua kontribusi utama, yaitu kontribusi metodologis dan kontribusi substantif. Secara

metodologis, penelitian ini menunjukkan bahwa pipeline analisis sentimen yang menggabungkan pengumpulan data media sosial, preprocessing yang mempertimbangkan konteks bahasa Twitter, pelabelan berbasis anotasi manusia, ekstraksi fitur TF-IDF, serta klasifikasi menggunakan algoritma Naïve Bayes, dapat digunakan secara cukup andal untuk menganalisis isu-isu sosial yang sensitif di konteks Indonesia. Evaluasi ini menegaskan bahwa validitas analisis sentimen pada isu kemanusiaan tidak hanya ditentukan oleh performa algoritma, tetapi juga oleh ketepatan proses pelabelan dan sensitivitas preprocessing terhadap makna kontekstual teks media sosial. Secara substantif, hasil penelitian mengungkap bahwa wacana publik masyarakat Indonesia di Twitter terkait pengungsi Rohingya didominasi oleh sentimen positif, yang merefleksikan nilai empati, solidaritas, dan kepedulian kemanusiaan yang masih kuat. Namun demikian, penelitian ini juga menemukan adanya kekhawatiran dan resistensi yang tercermin dalam sentimen negatif, terutama yang berkaitan dengan persepsi ketidakadilan sosial, kecemburuan ekonomi, serta pengaruh pemberitaan media dan disinformasi yang beredar di ruang digital. Temuan ini menunjukkan bahwa sentimen negatif tidak sepenuhnya menandakan penolakan terhadap prinsip kemanusiaan, melainkan lebih merupakan ekspresi kegelisahan sosial terhadap kebijakan dan dampak lokal dari penanganan pengungsi. Dengan demikian, penelitian ini memberikan pemetaan awal terhadap dinamika wacana publik mengenai pengungsi Rohingya di Indonesia, sekaligus menegaskan bahwa analisis sentimen media sosial dapat berfungsi sebagai instrumen pendukung bagi pembuat kebijakan untuk memahami opini publik secara lebih komprehensif. Temuan ini diharapkan dapat menjadi dasar bagi perumusan strategi komunikasi publik dan kebijakan yang lebih sensitif terhadap kekhawatiran masyarakat, tanpa mengabaikan nilai-nilai kemanusiaan yang menjadi fondasi respons Indonesia terhadap krisis pengungsi

References

- Fahrezi, M. F., & Permana, A. A. (2022). Analisis sentimen opini masyarakat pada media sosial Twitter terhadap organisasi Aksi Cepat Tanggap menggunakan Naïve Bayes Classifier. *Jurnal Ilmiah*, 11(2), 113–121.
- Gede Aditra Pradnyana, I. P., & Agustini, K. (2024). *Konsep dasar data mining*. Denpasar: Penerbit Akademik.
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis sentimen aplikasi Ruang Guru di Twitter menggunakan algoritma klasifikasi. *Jurnal Teknoinfo*, 14(2), 115–123. <https://doi.org/10.33365/jti.v14i2.679>
- Indah. (2022). *Analisis sentimen keberhasilan pemerintah menangani Covid-19 pada Twitter menggunakan metode Naïve Bayes*. Universitas Dinamika Bangsa.
- Indradipradana, R. K. (2023). Kebijakan luar negeri Indonesia: Studi kasus penerimaan pengungsi Rohingya asal Myanmar tahun 2020–2022. *Jurnal Hubungan Internasional*, 8(2), 211–236.
- Ismail, A. R., & Hakim, R. B. F. (2023). Implementasi lexicon-based untuk analisis sentimen dalam menentukan rekomendasi pantai di DI Yogyakarta berdasarkan data Twitter. *Emerging Statistics and Data Science Journal*, 1(1), 37–46. <https://doi.org/10.20885/esds.vol1.iss.1.art5>
- Kartika Sari, A., Irsyad, A., Aini, D. N., Islamiyah, & Ginting, S. E. (2024). Analisis sentimen Twitter menggunakan machine learning untuk identifikasi konten negatif. *Adopsi Teknologi dan Sistem Informasi (ATASI)*, 3(1), 64–73. <https://doi.org/10.30872/atasi.v3i1.1373>
- Khatami, M. K. (2024). *Analisis sentimen Twitter menggunakan Naïve Bayes dan Support Vector Machine terhadap KPU pada Pemilihan Umum Presiden 2024*. Universitas Islam Negeri Syarif Hidayatullah Jakarta.
- McCallum, A., & Nigam, K. (1998). A comparison of event models for Naïve Bayes text classification. *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization*, 41–48.
- Moy, L. Y., & Kusuma, A. J. (2022). Latar belakang Indonesia menerima pengungsi Rohingya pada tahun 2015 (analisis konstruktivis). *Jurnal Ilmu Hubungan Internasional*, 1, 60–75.
- Multi Fani, S., & Santoso, R. (2023). Penerapan text mining untuk melakukan clustering data tweet akun Blibli pada media sosial Twitter menggunakan K-Means. *Jurnal Media Informatika Budidarma*, 7(2), 583–593.
- Mustofa, Y. A. (2023). Analisis sentimen terhadap penggunaan aplikasi Shopee menggunakan algoritma Support Vector Machine. *Jurnal Sistem Informasi*, 5, 32–35.
- Putra, V., Agni, D., Kurniawan, R., & Wijaya, Y. A. (2024). Akurasi Naïve Bayes untuk analisis sentimen Twitter berdasarkan split data. *Journal of Computing Engineering, System and Science*, 9(1), 238–250.
- Putro, H. F., Vlandari, R. T., & Saptomo, W. L. (2020). Penerapan metode Naïve Bayes untuk klasifikasi pelanggan. *Jurnal Informatika*, 8(2), 45–52.
- Ramadhani, T., Hermawan, P., & Dzikrillah, A. R. (2024). Penerapan metode Naïve Bayes untuk analisis sentimen pada ulasan pengguna aplikasi ChatGPT di Google Play Store. *BITS: Jurnal Teknologi Informasi*, 6(1), 430–439. <https://doi.org/10.47065/bits.v6i1.5400>
- Rifaldi, D., Fadlil, A., & Herman. (2023). Teknik preprocessing pada text mining menggunakan data tweet “mental health”. *Decode: Jurnal Pendidikan Teknologi Informasi*, 3(2), 161–171. <https://doi.org/10.51454/decode.v3i2.131>

- Sopamena, & Abbas, C. (2023). Pengungsi Rohingya dan potensi konflik serta kemajemukan horizontal di Aceh. *Jurnal Caraka Prabhu*, 7(2), 85–115. <https://doi.org/10.36859/jcp.v7i2.1927>
- Suarna, N., & Prihartono, W. (2024). Analisis sentimen ulasan aplikasi Threads di Google Play Store menggunakan algoritma Naïve Bayes. *Jurnal Informatika*, 8(1), 967–974.
- Syafnidawaty. (2020). *Data mining*. Tangerang: Universitas Raharja.
- Syakur, A. (2021). Implementasi metode lexicon-based untuk analisis sentimen kebijakan pemerintah dalam pencegahan Covid-19 pada Twitter. *Jurnal Ilmiah Informatika Komputer*, 26(3), 247–260. <https://doi.org/10.35760/ik.2021.v26i3.4720>
- UNHCR. (2023). *Rohingya emergency overview*. United Nations High Commissioner for Refugees.
- Yuniar, E., Utsalinah, D. S., & Wahyuningsih, D. (2022). Implementasi scraping data untuk sentiment analysis pengguna dompet digital menggunakan algoritma machine learning. *Jurnal Janitra Informatika dan Sistem Informasi*, 2(1), 35–42. <https://doi.org/10.25008/janitra.v2i1.145>