

## Analisis Sentimen Masyarakat Terhadap Isu Migrasi Rohingya Ke Indonesia

Ulfa Kurniasih<sup>a</sup>, Akrim Teguh Suseno<sup>b</sup>

<sup>a</sup>Ekonomi Syariah, Ekonomi dan Bisnis Islam, UIN K.H. Abdurrahman Wahid, ulfa.kurniasih@uingusdur.ac.id

<sup>b</sup>Informatika, Sains dan Teknologi, ITSNU Pekalongan, [akrim.teguh@gmail.com](mailto:akrim.teguh@gmail.com)

Submitted: 11-01-2025, Reviewed: 20-01-2024, Accepted 27-01-2025

<https://doi.org/10.47233/jteksis.v7i1.1815>

### Abstract

*The phenomenon of the Rohingya refugee exodus is something that attracts a lot of public interest on social media. The public's views that reap the pros and cons of the arrival of Rohingya refugees have led to debates with narratives of hatred for each other with attitudes of rejection or some accepting Rohingya refugees on social media, especially X on the basis of human rights. From this problem, it is necessary to map public sentiment which can later help the government in determining policies to handle the arrival of Rohingya refugees and minimize unrest or debate in the community. The data used in this study were tweets from users with the keyword "rohingya" which were analyzed using the text mining method with the Decision tree 5.0 and Naive Bayes algorithms. The results showed that negative sentiment was still dominant on social media X with an average of 49.5%, positive sentiment 18.5% and neutral sentiment 27%. For validity testing using the K-Fold Validation method, the test showed that the Naïve Bayes algorithm had a better accuracy level with an accuracy level of 83% while the decision tree C.50 only had an accuracy level of 78%. The results of the study also illustrate that Indonesian society through social media X still has a negative view or rejects the arrival of Rohingya refugees in Indonesia.*

**Keywords:** sentiment analysis, refugees, rohingya, naive bayes, decision tree

### Abstrak

Fenomena eksodus pengungsi Rohingya merupakan hal yang menarik banyak minat masyarakat di media sosial. Pandangan masyarakat yang menuai pro dan kontra terhadap kedatangan pengungsi rohingya yang menyebabkan munculnya perdebatan dengan narasi kebencian satu sama lain dengan adanya sikap penolakan ataupun ada yang menerima pengungsi rohingya di media sosial khususnya X atas dasar hak asasi manusia. Dari permasalahan tersebut perlu dilakukan pemetaan terhadap sentimen masyarakat yang nantinya dapat membantu pemerintah dalam menentukan kebijakan untuk menangani kedatangan pengungsi rohingya dan meminimalisir adanya keributan atau perdebatan dilingkungan masyarakat. Data yang digunakan pada penelitian ini adalah tweets dari para pengguna dengan keyword "rohingya" yang dianalisis dengan metode *text mining* dengan algoritma *Decision tree* 5.0 dan *Naive bayes*. Hasil penelitian menunjukkan bahwa sentimen negatif masih dominan di sosial media X dengan rata-rata 49,5%, sentimen positif 18,5% dan sentimen netral 27%. Untuk pengujian validitas menggunakan metode *K-Fold Validation*, pengujian tersebut menunjukkan bahwa algoritma *Naive Bayes* mempunyai tingkat akurasi yang lebih baik dengan tingkat akurasi sebesar 83% sedangkan *decision tree* C.50 hanya memiliki tingkat akurasi sebesar 78%. Hasil penelitian tersebut juga menggambarkan bahwa masyarakat indonesia melalui media soisial X masih berpadangan negatif atau menolak kedatangan pengungsi rohingya di indonesia.

**Keywords:** sentimen analisis, pengungsi, rohingya, naive bayes, decision tree

*This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license*



### PENDAHULUAN

Salah satu permasalahan pengungsi yang saat ini sedang menjadi perhatian dunia ialah pengungsi asal Myanmar atau yang dikenal dengan etnis Rohingya. Dalam sejarah Myanmar, hubungan antara pemerintah pusat dengan etnis minoritas khususnya muslim adalah hubungan antara penguasa dan pemerontok[1]. Terlebih adanya kampanye imigran ilegal yang mulai diproklamkan pada tahun 1978 menyebabkan kondisi Rohingya semakin terpojok hingga pecahnya konflik di Myanmar pada tahun 2012 yang menewaskan ratusan jiwa dan menyebabkan sekitar 140.000 orang Rohingya melarikan diri ke luar negeri.

Data UNHCR menyebutkan bahwa warga Rohingya tidak hanya mencari keamanan di

Indonesia melainkan juga Bangladesh (>960.000), Malaysia (>107.000), dan India (>22.000). Data terakhir menyebutkan terdapat 1.445 pengungsi Rohingya yang mendarat di sejumlah pantai di Aceh. Sebagian besar pengungsi terdiri atas perempuan dan anak-anak (lebih dari 70%), serta kelompok rentan lainnya seperti penyandang disabilitas dan lansia.

Kehadiran gelombang pengungsi etnis Rohingya yang mencari suaka ke wilayah Asia Tenggara rupanya menimbulkan polemik tersendiri bagi masyarakat Indonesia, terutama masyarakat Aceh. Kedatangan secara besar-besaran hingga total pengungsi mencapai lebih dari 1400 di Desember 2023 [2] membuat masyarakat dan pemerintah harus berbagi sandang dan pangan kepada para pengungsi

padahal ekonomi masyarakat di sekitar masih membutuhkan bantuan. Selain itu, hadirnya berbagai isu yang tersebar di media sosial terkait kemampuan adaptasi para pengungsi yang masih kurang terhadap adat dan budaya di Aceh membuat masyarakat menolak kehadiran mereka.

Opini masyarakat yang berkembang di media sosial menjadi semakin beragam sejak munculnya pernyataan Wakil Presiden, Ma'aruf Amin yang menyebut Pulau Galang sebagai lokasi yang dapat dipertimbangkan sebagai lokasi pengungsi Rohingya. Alasannya, Pasal 28 UUD 1945 menyebutkan bahwa Indonesia mengakui hak para pencari suaka bagi semua orang dan Indonesia merupakan salah satu negara yang mengakui Deklarasi Universal Hak Asasi Manusia. Namun, pada sisi lain Indonesia tidak termasuk negara pihak Konvensi Jenewa 1951 sehingga pada dasarnya Indonesia tidak memiliki kewajiban untuk menerima pengungsi yang masuk[3]. Hal ini memperkeruh sentimen masyarakat terkait pro kontra kehadiran para pengungsi. Ragam narasi dukungan maupun narasi penolakan dan kebencian mengenai Rohingya bermunculan di media sosial X dan menjadi isu nasional yang ramai diperbincangkan. Tagar "Rohingnya" sempat menduduki trending topik nomor 1 di X pada awal bulan Desember 2023.

Artikel yang tulis oleh Albert dan Maizland yang berjudul "The Rohingya Crisis" menjelaskan bahwa sejak 2017 berbagai daerah telah mengecam penganiayaan terhadap Rohingya seperti di kota-kota Pakistan, India, Thailand, Indonesia, dan Bangladesh. Masyarakat Rohingya telah mengungsi di berbagai daerah di Asia khususnya Bangladesh, Malaysia, Indonesia dan India. Namun telah terjadi penolakan oleh masyarakat sekitar di negara masing-masing hingga pada tahun 2019 negara-negara pengungsi sepakat memulangkan pengungsi rohingya ke Myanmar tetapi pemimpin komunitas Rohingya menolak karena tidak ada jaminan keamanan di Myanmar. Polemik tersebut terus terjadi namun tidak memiliki solusi yang pasti terhadap keadaan pengungsi Rohingya. Negara-negara di Asia Tenggara umumnya tidak memiliki kerangka hukum untuk melindungi hak-hak pengungsi, dan sepuluh negara anggota Perhimpunan Bangsa-Bangsa Asia Tenggara (ASEAN) belum mengoordinasikan respons terhadap krisis yang semakin parah. Indonesia, Malaysia, Myanmar, dan Thailand—semuanya anggota ASEAN—belum meratifikasi konvensi pengungsi PBB atau protokolnya. ASEAN sendiri sebagian besar bungkam mengenai penderitaan warga Rohingya dan meningkatnya jumlah pencari suaka di negara-negara anggotanya, terutama karena komitmen para anggotanya terhadap prinsip non-intervensi dalam setiap urusan negara[4].

Pada penelitian ini untuk melihat sentimen masyarakat terhadap komunitas rohingya di Indonesia menggunakan metode analisis sentimen yang mengklasifikasikan opini, emosi, atau sentimen dalam teks berdasarkan orientasi emosional, baik positif maupun negatif[5][6]. Sentimen analisis merupakan bagian dari text mining[7] bertujuan mengubah teks menjadi bentuk yang dapat diolah dan memberikan wawasan yang berarti [8]. *Text mining* dapat digunakan untuk analisis pada kategori klasifikasi, *information retrieval*, dan *information extraction*. *Text mining* bertujuan untuk mengklasifikasikan dokumen dalam bentuk teks menjadi *insight* yang bermanfaat untuk organisasi baik pemerintah maupun swasta. Pemanfaatan *text mining* telah banyak dimanfaatkan di berbagai bidang seperti sentimen analisis pada kasus covid-19, isu agama, dan deteksi sarkasme [9], [10], [11]. Dalam menerapkan *text mining*, diperlukan langkah analisis yang bernama *Text Preprocessing*. *Text Preprocessing* merupakan serangkaian langkah yang dilakukan pada data teks untuk membersihkan dan mengubahnya menjadi format yang lebih sesuai untuk analisis. Langkah-langkah ini bertujuan untuk meningkatkan kualitas data, menghilangkan *noise*, dan mempersiapkannya agar dapat diolah oleh algoritma pemrosesan *text mining* [12][13].

Pada penelitian terdahulu yang dilakukan oleh Aziza & Rani melakukan penelitian terkait konten dakwah radikal Indonesia menggunakan analisis sentimen dan text mining melalui media sosial X. Penelitian tersebut melihat bagaimana kelompok masyarakat dalam bentuk individual maupun kelompok menyebarkan ide ekstremis maupun sikap kemarahan dalam bentuk rasial. Penelitian ini mengklasifikasikan sentimen masyarakat terhadap konten-konten radikalisme. Metode analisis data yang digunakan adalah teknik klasifikasi dengan algoritma K-Nearest Neighbor (K-NN) dengan hasil akurasi sebesar 66,37% [14]. Namun pada penelitian ini tidak memberikan gambaran hasil sentimen positif maupun negatif sehingga tidak bisa dilihat kecenderungan masyarakat terhadap isu radikalisme konten dakwah radikal di Indonesia. Perbedaan yang dilakukan penulis dan Aziza & Rani terletak pada objek, pengembangan teknik analisis data dan hasil yang diperoleh. Pada penelitian yang dilakukan penulis dataset yang digunakan yaitu pengungsi Rohingya di Indonesia tahun 2023 kemudian teknik analisis data peneliti menggunakan algoritma decision tree C5.0 & *naive bayes* yang memiliki perbedaan perhitungan dalam analisis data dengan K-NN. Hasil dari penelitian penulis akan memberikan gambaran sentimen positif, negatif dan netral dari masyarakat melalui sosial media X terhadap pengungsi Rohingya di Indonesia.

Penelitian lainnya dilakukan oleh Chowdhury et al. (2018) yang meneliti tentang analisis sentimen pada komentar di media sosial Facebook tentang perpindahan orang-orang Rohingya di Bangladesh. Penelitian ini menggunakan algoritma *Naive bayes* dan *Support Vector Machine (SVM)* dalam menganalisis data kalimat yang terdapat dalam komentar di media sosial Facebook. Hasil analisis tersebut memberikan tingkat akurasi sebesar 78% pada *naive bayes* dan 81% pada SVM. Algoritma SVM memiliki tingkat akurasi yang cukup baik dalam beberapa penelitian lain [15][16]. Namun, penelitian ini mengabaikan posisi kata dalam kalimat sehingga kesulitan dalam menetapkan akurasi konteks dalam suatu postingan kalimat. Selain itu penelitian ini tidak menjelaskan hasil sentimen positif, negatif maupun netral terhadap pengungsi Rohingya di Bangladesh melalui media sosial Facebook [17]. Perbedaan penelitian ini dengan penelitian Chowdhury et al. (2018) terletak pada objek penelitian, kebaruan teknik analisis data hasil sentimen menyarakat. Penulis menggunakan dataset berupa komentar di X terhadap kedatangan pengungsi Rohingya di Indonesia tahun 2023 dan penggunaan algoritma *decision tree C5.0* yang lebih terbaru dibandingkan dengan SVM serta hasil sentimen positif, negatif dan netral akan dijabarkan sehingga dapat membantu menjadi bahan pertimbangan pemerintah dalam menyikapi para pengungsi Rohingya di Indonesia.

Rochmawati & Wibawa (2018) melakukan penelitian terhadap sentimen masyarakat tentang konflik antara penduduk Rakhine dengan tentara Myanmar menggunakan *Naive bayes*. Penelitian tersebut mendapatkan kesimpulan bahwa sentimen masyarakat terhadap kelompok yang pro dan kelompok yang kontra seimbang. Sedangkan kelompok netral jumlahnya lebih sedikit. Temuan penelitian yang menggunakan data dari cuitan di X tersebut ialah banyak teks yang tidak terdeteksi jenis emosi lainnya karena kesulitan dalam mengolah data teks. Hal ini mungkin disebabkan oleh karena kurangnya penguasaan terhadap tata bahasa sebab data yang diperoleh merupakan data yang menggunakan bahasa asing [18]. Penelitian Rochmawati & Wibawa (2018) memiliki perbedaan dengan penelitian ini. Perbedaan tersebut salah satunya terletak pada objek penelitian. Namun algoritma *naive bayes* memiliki tingkat akurasi yang baik pada penelitian-penelitian lainnya [15] sehingga menjadi pertimbangan dalam pendekatan penelitian yang akan dilakukan oleh penulis

Penelitian yang dilakukan oleh penulis menggunakan objek cuitan masyarakat Indonesia yang terdapat di X. Cuitan tersebut berupa respons masyarakat terhadap kedatangan pengungsi Rohingya yang semakin bertambah di Indonesia, khususnya pada tahun 2023. Selain itu, penelitian ini

nantinya menggunakan algoritma *Decision tree C5.0* dan *Naive bayes* untuk melihat perbandingan akurasi yang lebih baik dalam model analisis data. Penggunaan algoritma *decision tree* karena memiliki kelebihan pada tingkat akurasi yang lebih baik pada algoritma lainnya [19].

Kebaharuan (novelty) pada penelitian ini yaitu pada algoritma yang digunakan untuk analisis. Penelitian yang dilakukan oleh Chowdhury et al. (2018) menggunakan algoritma *Naive bayes* dan *Support Vector Machine (SVM)* dengan objek penelitian Facebook. Kemudian pada penelitian yang dilakukan oleh Rochmawati & Wibawa (2018) hanya menggunakan algoritma *Naive bayes* dengan objek penelitian X. Pada penelitian yang dilakukan oleh penulis ini menggunakan algoritma *Naive bayes* dan *Decision tree 5.0* dengan objek penelitian X. Penelitian ini belum ditemukan adanya studi kasus yang serupa dengan penggunaan 2 algoritma tersebut di objek penelitian X. Berdasarkan permasalahan di atas, penelitian ini menggunakan sentimen analisis untuk melihat kritik, pendapat, sikap, emosi, dan respons masyarakat tentang keberadaan X di Indonesia menggunakan X sebagai dataset.

## METODE PENELITIAN

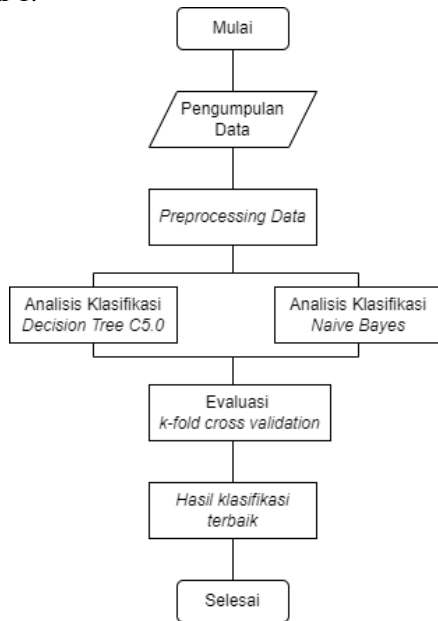
Dengan menggunakan metode campuran (mixed method), penelitian ini menggabungkan antara metode kualitatif dan kuantitatif secara bersamaan. Pada metode kualitatif menggunakan data teks dari tweet maupun komentar dari pengguna di media sosial X yang berbentuk sebuah kalimat maupun paragraf. Kemudian pada metode kuantitatif, data teks dari X tersebut ditransformasikan menjadi numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Metode ini digunakan dalam pengolahan teks dan pencarian informasi untuk menilai pentingnya suatu kata dalam dokumen relatif terhadap koleksi dokumen lainnya [20].

Penelitian ini menggunakan pendekatan dengan menggunakan metode text mining untuk dapat menganalisis sentimen publik. Tujuannya adalah untuk memahami bagaimana gambaran maupun sentimen publik terhadap suatu topik yang sedang dibicarakan di media sosial—dalam hal ini Rohingya—X. Sikap, pendapat, penilaian, dan emosi yang hadir dalam bentuk tulisan tersebut dapat menggambarkan pandangan masyarakat pengguna media sosial secara positif, negatif, maupun netral. Populasi pada penelitian adalah data teks dari X atau disebut juga cuitan yang menggunakan kata “Rohingya”. Data pada penelitian ini dikumpulkan dengan pendekatan kuantitatif. Teknik pengambilan data pada penelitian ini menggunakan teknik random sampling dengan jumlah 1000 postingan kata/kalimat di X yang

mengandung kata “Rohingya” pada rentang waktu November 2023 hingga Maret 2024.

### 2.1 Tahapan Penelitian

Pada tahapan penelitian dapat ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pada gambar 1 menunjukkan tahapan penelitian yang dimulai dari pengumpulan data kemudian dilakukan validitas sekaligus screening data tweet melalui preprocessing data. Langkah selanjutnya adalah analisis data menggunakan klasifikasi *Decision tree* C5.0 dan *Naive bayes* kemudian dilakukan evaluasi hasil menggunakan metode K-Fold Cross Validation sehingga dapat diperoleh hasil klasifikasi terbaik terhadap sentimen masyarakat terhadap pengungsi Rohingya di Indonesia.

## HASIL DAN PEMBAHASAN

Penelitian ini melakukan pengumpulan data menggunakan teknik crawling data media sosial khususnya X. Pengambilan data X bersumber dari status maupun komentar yang mengandung kata “Rohingya” dan batasan tweet hanya di wilayah Indonesia dengan jumlah 1000 tweet. Data yang diambil berupa username, teks, dan date, dengan rentang waktu dari 1 November 2023 hingga 31 Maret 2024. Berikut adalah contoh data yang berhasil diambil pada X yang ditunjukkan pada tabel 1:

No	Nama Pengguna	Tweet
1	ribbtions	temen gue beneran benci bgt sama etnis rohingya...-

2	PutriSyalala8	@siapayaa5555 @feblastiawan @SmudgeStan fasis gmn sih, mosok menolak datangnya rohingya dikatain fasis...luar biasa lu..
3	lunaatmo	@lautaromarco_ @kaoriyuuki_ @rif_doel @convomfs I see.. kasian ya. Siapa juga yang milih lahir jadi Rohingya. Kalau bisa milih, jadi anak konglomerat aja.
4	susiloian5	@Kratozzss @babybohoyoho @kyuisann Apa urusan rohingya sama pak anies?
5	PutriSyalala8	@siapayaa5555 @feblastiawan @SmudgeStan sangar2 juga buzzer rohingya...wao
6	ludosiso	@tanyakanrl Tp kalo ngusir pengungsi Rohingya dg cara anarkis dan nirempati sih kontra.
7	Samsurijal_A	Etnis Rohingya di Rakhine, Dari Entitas Hingga Kehilangan Identitas <a href="https://t.co/pj0CZKmUp5">https://t.co/pj0CZKmUp5</a>
8	Ormarsh_Putih	@AdiSapu97789863 @joaydwj -&gt; Apa hubungannya pengangguran sama Rohingya? Gak nyambung ih...
9	PutriSyalala8	@merseysideliv @feblastiawan @SmudgeStan sebenarnya yg jadi masalah utama itu bukan soal kelakuan rohingya baik ato buruk, tp soal datangnya mereka ke indo...itu masalahnya..
10	nickkaskus	@tikaalmira Kamu tuh, bela rohingya kaya ke saudara kandung. Apakah kamu sudah turun kelapangan ngurusin orang rohingya?

Pada data di atas menunjukkan bahwa data tersebut terdiri dari 3 atribut yaitu tanggal posting, username, dan tweet yang berupa status ataupun komentar terhadap status pengguna lainnya. Data tersebut masih bersifat kotor karena ada beberapa atribut yang tidak digunakan, kata tidak baku, terdapat tanda baca, adanya link yang tidak perlu dan lain-lain sehingga perlu dilakukan pembersihan data melalui preprocessing data.

### 3.1 Pembahasan

#### A. Preprocessing Data

Pada tahap ini melalui 5 proses yaitu *cleaning data*, *case folding*, *tokenizing*, *stemming* dan *stopword removal*. Berikut adalah hasil dari preprocessing data yang dapat ditunjukkan pada Tabel 2:

Tahapan	Hasil
Kalimat dasar	Lha itu beneran rohingya atau migran cari kerja..? Kalau migran cari kerja berkedok pengungsi rohingya ya mending di genosida saja!
Data Cleaning	Lha itu beneran rohingya atau migran cari kerja, Kalau migran cari kerja berkedok pengungsi rohingya mending di genosida saja
Case folding	itu rohingya atau migran cari kerja, kalau migran cari kerja berkedok pengungsi rohingya mending di genosida saja

Tokenizing	'itu' 'rohingya' 'atau' 'migran' 'cari' 'kerja' 'kalau' 'migran' 'cari' 'kerja' 'berkedok' 'pengungsi' 'rohingya' 'mending' 'di' 'genosida' 'saja'
Stemming	'itu' 'rohingya' 'atau' 'migran' 'cari' 'kerja' 'kalau' 'migran' 'cari' 'kerja' 'kedok' 'pengungsi' 'rohingya' 'mending' 'di' 'genosida' 'saja'
Stopword removal	'rohingya' 'imigran' 'cari' 'kerja' 'imigran' 'cari' 'kerja' 'kedok' 'pengungsi' 'rohingya' 'genosida'

Pada tabel di atas menjelaskan bahwa terdapat 6 fase dalam proses preprocessing data. Pada fase pertama yaitu data cleaning yang menghapus atribut yang tidak dibutuhkan seperti date dan username sehingga yang tersisa hanya data teks. Kemudian pada data teks tersebut hanya diambil yang berbahasa Indonesia, menghapus duplikasi data, dan menghapus teks yang tidak bermakna sehingga data yang terkumpul yaitu sebesar 658 data. Data cleaning juga menghapus tanda baca serta karakter-karakter simbol pada kalimat sehingga emoticon dan lainnya tidak digunakan dalam penelitian ini. Pada fase kedua yaitu case folding yang mengubah semua huruf dalam teks menjadi huruf kecil. Hal tersebut bertujuan memastikan konsistensi dalam representasi teks, seperti pada kata “Kalau” pada fase data cleaning akan menjadi “kalau” pada fase case folding.

Tahap selanjutnya adalah tokenizing yang berupa token-token atau bagian tertentu yang terbentuk berdasarkan kalimat maupun paragraf yang dipecah. Pemisah antar token dapat berupa spasi dan tanda baca. Contoh Tokenisasi: “Etnis Rohingya di Rakhine”, menghasilkan 4 token yaitu “etnis”, “rohingya”, “di”, “rakhine”. Token-token ini yang akan di proses untuk analisis lebih lanjut. Pada tahap ini, teks dibagi menjadi bagian-bagian yang lebih kecil yang disebut “token”. Token dapat berupa kata, frasa, simbol, atau entitas lainnya yang relevan dengan analisis yang dilakukan. Proses tokenisasi sangat penting dalam data mining karena membantu mengurai teks menjadi unit yang lebih kecil dan dapat diolah lebih lanjut. Tahap ini biasanya dilakukan dengan cara melakukan pemisahan berdasarkan spasi.

Pada proses stemming adalah menghapus akhiran dari kata untuk mengubahnya ke bentuk dasar atau akar. Seperti pada kata “berkedok” maka akan diubah ke bentuk dasar yaitu “kedok”. Pada tahap selanjutnya yaitu *stopword removal* yang bertujuan menghapus kata-kata yang dianggap tidak memiliki makna penting atau relevansi dalam analisis. Pada analisis ini diperlukan database untuk sebagai acuan dalam kata dasar yang relevan dalam analisis. Berikut adalah beberapa contoh pada data *stopword* yang digunakan dalam analisis ini

Tabel 3. Contoh data *stopword*

Kata Dasar
------------

ada	bagai
adalah	bagaikan
adanya	bagaimana
adapun	bagaimanakah
agak	bagaimanapun
agaknya	bagi
agar	bagian

Hasil dari analisis tersebut kemudian dibagi menjadi data training dan data testing. Data training digunakan untuk membentuk pola yang kemudian di uji dengan data testing untuk melihat tingkat akurasi dari algoritma klasifikasi *Naive bayes* maupun *decision tree C5.0* sehingga dapat memberikan rekomendasi algoritma yang digunakan untuk prediksi sentimen analisis pada kasus rohingya. Dataset yang digunakan yaitu data training dan data testing diperoleh dari pembagian 658 data dengan perbandingan 50:50 sehingga data training dan data testing masing-masing berjumlah 329 data. Dataset merupakan data yang sudah berisi sentimen positif, negatif dan netral pada setiap text secara manual berdasarkan sentimen pada kalimat teks masing-masing.

#### B. Analisis Klasifikasi *Naive bayes*

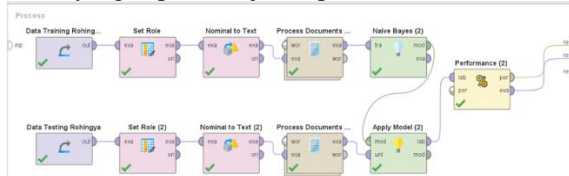
Penelitian ini menggunakan algoritma *Naive bayes* dan tools bantuan yang digunakan yaitu rapidminer. Data yang digunakan yaitu 658 data dari hasil preprocessing data yang dibagi menjadi 2 yaitu data training dan data testing. Data training berjumlah 329 dengan rincian label sentimen 157 negatif (48%), 101 netral (31%) dan 71 positif (22%) yang contohnya dapat ditunjukkan pada tabel berikut:

Tabel 4. Data Training

No	Text	Sentimen
1	@abheemanyoen @YurisAbrori @rgantas Eh idiot sebelum comment pahami dulu psikologi social ya dongo. Yg jadi masalah rohingya memang beberapa kali ada masalah kan, kayak berak di tambak warga.	negatif
2	@achehsultanate Mknya sejak awal yg ngemeng nolak Rohingya demi bela rakyat Aceh itu bullshit bgt. Plg di waktu lain dia hina dina penegakan syari'at di Aceh.	negatif
3	@Addidiero_ @convomf True. Karena media luar nge goreng nya warga rohingya yg 'tertindas', in fact mereka yg 'jahat'. Manipulatif.	negatif
4	@BBCIndonesia Dulu ngejek orang Malaysia, katanya tiada perikemanusiaan. Usir Rohingya. Terbukti sekarang	negatif
5	@armyteamunivers @siapayaa5555 @asiangames2023 @afifahafra79 Lah, lisan lu lebih kotor ajii gilak masih sama sma manusia aja kek gitu.. keknya doa dia yg dijabah apalgi doa org teraniaya kek rohingya lebih mudah dijabah	negatif
6	@___4kim @efanhamza Nak kaitkan agama plak, nk suruh hantar dorg balik ke tempat asal je bukan nk anaiya dorang pun, tengok Palestine berjuang nk dapat tanah air dia balik, asal x rohingya?	netral
7	@_kamilmuhammad1 Wah refugee afghan dan rohingya diurus dulu dah baru omong Palestine	netral

8	@_shfmhrn @sir_amirsyarif @margianta Maaf, mas. Benar bahwa warga rohingya tidak seburuk itu. Tapi disini yang saya lihat, warga Aceh tidak ingin menerima lagi karena takut mengulangi kesalahan yg sama akibat oknum pengacau dari Rohingya yg meresahkan Aceh.	netral
9	@Aceh @Refugees @palangmerah @BNPB_Indonesia @humasaceh @DPR_Aceh @UN @UNmigration Kalau dibawa ke Jakarta atau Bali, waduh, suka Rohingya. Lagian kalau sekali dibagi KTP. Mereka bisa ke mana mana di Indonesia, bisa berniaga. Lagian dibagi makanan gratis, penampungan di kondo, gratis, mobil gratis, medikal gratis. Pasti mereka bersyukur.	positif
10	@AdiSapu97789863 @joaydwj -&gt; Apa hubungannya pengangguran sama Rohingya? Gak nyambung ih... Emang lowongan kerja ellu di ambil sama Rohingya ya? -&gt; Kenapa Rohingya datang baru ellu peduli kemiskinan di Indonesia? Bukannya diatas itu masuk program pemerintah ya? Kok semua salah pengungsi? Logikanya Oleng	netral

Pelabelan sentimen pada data training menggunakan metode manual yaitu membaca dari kalimat teks setiap tweet kemudian diberikan label sentimen. Sedangkan untuk data testing berjumlah 329 yang belum memiliki label negatif, netral maupun positif. Data training dianalisis menggunakan algoritma Naïve Bayes sehingga membentuk sebuah pola dan kemudian pola tersebut diujikan ke dalam data testing sehingga menghasilkan prediksi sentimen yang berlabel negatif, positif dan netral. Berikut adalah proses dari analisis *Naive bayes* dengan bantuan tools rapid miner yang dapat ditunjukkan pada Gambar 2



Gambar 2. Proses analisis Naïve Bayes

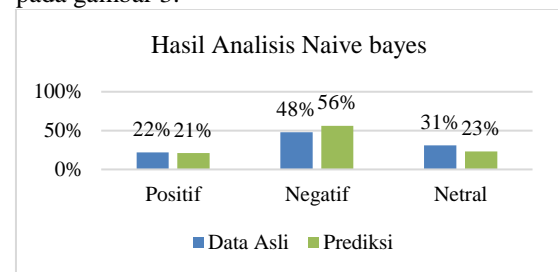
Pada gambar diatas menunjukkan bahwa data training diambil terlebih dahulu kemudian dilakukan set role untuk menentukan label atau target yang dituju. Apabila melihat dari Tabel 3 maka kolom sentimen merupakan target. Kemudian dilakukan data transformasi dari nominal to text untuk mengkonversi apabila ada karakter nominal di dalam data training diubah menjadi bentuk string (teks). Langkah selanjutnya yaitu process document yang didalamnya terdapat proses Case folding, Tokenizing, Stemming dan *Stopword removal*. Setelah itu data telah siap untuk dilakukan pembentukan pola melalui algoritma Naïve Bayes. Setelah pola terbentuk maka akan diujicobakan dengan data testing melalui operator apply model. Data testing yang melalui proses yang sama seperti data training dari set role hingga process document. Hasil dari ujicoba tersebut juga dilakukan evaluasi dengan *k-fold cross validation* melalui operator

performance. Berikut adalah contoh hasil analisis dari perhitungan menggunakan algoritma *naive bayes* yang ditunjukkan pada Tabel 5.

Tabel 5. Data Training

No.	Sentimen	Prediction	Confidence (negatif)
1	negatif	negatif	1
2	negatif	negatif	1
3	negatif	negatif	1
4	negatif	negatif	1
5	negatif	negatif	1
6	negatif	negatif	1
7	negatif	negatif	1
8	negatif	negatif	1
9	negatif	negatif	1
10	negatif	negatif	1
11	negatif	negatif	1
12	negatif	negatif	1

Pada Tabel 5 menunjukkan bahwa antara sentimen yang diisi secara manual dengan prediksi sentimen dari algoritma Naïve Bayes menunjukkan banyak kesamaan. Hasil analisis dengan menggunakan algoritma Naïve Bayes ditunjukkan pada gambar 3.



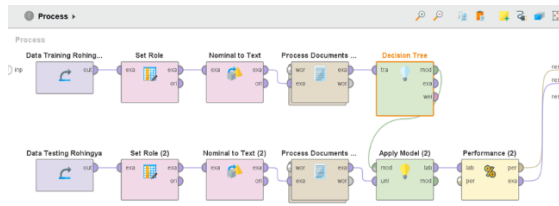
Gambar 3. Hasil prediksi analisis Naïve Bayes

Pada hasil analisis di gambar 3 menunjukkan bahwa data prediksi memiliki tingkat nilai yang hampir sama seperti pada nilai negatif, data asli menunjukkan 48% sedangkan data prediksi menunjukkan 56% (184), untuk sentimen netral pada data asli menunjukkan 31% sedangkan pada data prediksi menunjukkan 23% (75). Sedangkan untuk sentimen positif menunjukkan bahwa nilai asli sebesar 22% dan nilai prediksi sebesar 21% (70).

### C. Analisis Klasifikasi *Decision tree* C5.0

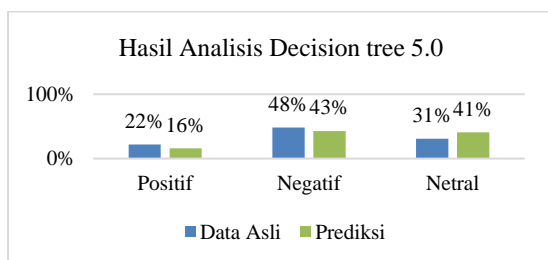
Analisis selanjutnya yaitu menggunakan algoritma *Decision tree* C5.0 dan tools bantuan yang digunakan yaitu rapidminer. Data yang digunakan sama dengan analisis Naïve Bayes yaitu 658 data dari hasil preprocessing data yang dibagi menjadi data training dan data testing. Hasil analisis data training yang berjumlah 329 dengan rincian label sentimen sebagai berikut 157 negatif (48%), 101 netral (31%) dan 71 positif (22%). Sedangkan untuk data testing berjumlah 329 yang belum memiliki label. Data training dianalisis menggunakan algoritma *Decision tree* sehingga membentuk sebuah pola dan kemudian pola tersebut diujikan ke dalam data testing sehingga menghasilkan prediksi

sentimen yang berlabel negatif, positif dan netral. Berikut adalah proses analisis data algoritma *decision tree* menggunakan tools rapidminer yang ditunjukkan pada gambar 4:



Gambar 4. Analisis tools rapidminer *Decision tree*

Pada gambar 4 menggambarkan alur dari proses analisis *decision tree* yang hampir sama dengan proses dari analisis menggunakan algoritma Naïve Bayes. Analisis dimulai dari menginputkan data training. Data tersebut ditentukan berdasarkan target kolom yang akan dianalisis, yakni kolom sentimen seperti yang ditunjukkan pada Tabel 4. Kemudian data yang terdapat angka nominal diubah menjadi bentuk teks. Setelah itu akan dilakukan process document yang di dalamnya terdapat proses Case folding, Tokenizing, Stemming dan Stopword removal. Proses selanjutnya dilakukan pembentukan pola melalui algoritma *decision tree*. Setelah pola terbentuk maka akan diujicobakan dengan data testing melalui operator apply model. Data testing yang melalui proses yang sama seperti data training dari set role hingga process document. Hasil dari ujicoba tersebut juga dilakukan evaluasi dengan k-fold cross validation melalui operator performance. Hasil dari analisis tersebut dapat ditunjukkan pada Gambar 5.



Gambar 5. Hasil prediksi analisis Naïve Bayes

Pada hasil analisis di gambar 5 menunjukkan bahwa data prediksi memiliki tingkat nilai yang hampir sama seperti pada nilai negatif, data asli menunjukkan 48% sedangkan data prediksi menunjukkan 43% (140), untuk sentimen netral pada data asli menunjukkan 31% sedangkan pada data prediksi menunjukkan 41% (135). Sedangkan untuk sentimen positif menunjukkan bahwa nilai asli sebesar 22% dan nilai prediksi sebesar 16% (54).

#### D. Evaluasi *K-Fold Cross Validation*

Teknik ini dilakukan untuk mengukur performa model secara lebih akurat dan reliabel,

yakni dengan melihat hasil dari klasifikasi yang dihasilkan dari perhitungan manual dengan hasil dari klasifikasi data dengan *decision tree* yang mengimplementasikan algoritma Naïve Bayes dan *Decision tree* C5.0. Evaluasi dengan *K-Fold Cross Validation* nilai k yaitu 3 yang menggambarkan jumlah fold data dan setiap nilai subset mempunyai data dengan nilai yang berbeda. Iterasi pada pengujian yang akan dilakukan yaitu sebanyak tiga kali. Berikut adalah hasil dari Iterasi pengujian pada algoritma *Naive bayes* dan *Decision tree* C5.0 ditunjukkan pada Tabel 5 dan Tabel 6.

Tabel 5. Iterasi Pengujian

Iterasi	Data Testing	Data Training
1	F1	F2, F3
2	F2	F1, F3
3	F3	F1, F2

Penelitian ini melakukan uji validasi sebanyak tiga kali dengan rata-rata nilai error pada data uji yaitu 17%. Hasil iterasi pengujian ditunjukkan pada Tabel 6.

Tabel 6. Hasil Iterasi Pengujian Naïve Bayes dan *Decision tree*

Data yang diuji	Nilai Error	
	<i>Naive bayes</i>	<i>Decision tree</i>
Iterasi ke-1	16%	20%
Iterasi ke-2	18%	24%
Iterasi ke-3	17%	21%
Nilai rata-rata error	17%	22%

Berdasarkan hasil uji validasi, nilai rata-rata error pada algoritma *Naive bayes* adalah 17% dengan nilai akurasi 83%. Sementara itu, pada algoritma *Decision tree* C5.0 yang memiliki tingkat error yaitu 22% dengan tingkat akurasi yaitu 78%.

#### E. Rekomendasi

Hasil dari analisis sentimen masyarakat dengan metode komputasional text mining dengan algoritma naïve bayes maupun *decision tree* 5.0 dengan kata kunci “rohingya” pada sosial media X menunjukkan bahwa sentimen negatif masih dominan dengan nilai rata-rata 49,5%, sentimen positif 18,5% dan sentimen netral 27%. Selain itu hasil evaluasi dengan menggunakan *K-Fold Cross Validation* menunjukkan bahwa algoritma Naïve Bayes memiliki tingkat akurasi rata-rata yang lebih baik yaitu 83% dengan tingkat error sebesar 17% sedangkan algoritma *Decision tree* C5.0 memiliki tingkat akurasi rata-rata yang lebih rendah yaitu 78% dengan tingkat error sebesar 22%. Oleh sebab itu rekomendasi algoritma yang dapat digunakan untuk memprediksi sentimen masyarakat melalui media sosial X yaitu algoritma naïve bayes.

#### SIMPULAN

Berdasarkan latar belakang serta hasil analisis pada penelitian ini, dapat disimpulkan bahwa:

1. Analisis sentimen masyarakat dengan metode komputasional *text mining* dengan algoritma *naïve bayes* maupun *decision tree 5.0* dengan kata kunci “rohingya” pada sosial media X menunjukkan bahwa sentimen negatif masih dominan dengan rata-rata 49,5%, sentimen positif 18,5% dan sentimen netral 27%. Hal tersebut menunjukkan bahwa masyarakat indonesia melalui sosial media X masih cenderung memberikan sikap yang negatif terhadap kehadiran dari penduduk Rohingya di Indonesia.
2. Sentimen masyarakat menggunakan metode komputasional *text mining* dengan algoritma *Naïve bayes* pada kata ”Rohingya” di sosial media X menunjukkan bahwa sentimen positif sebesar 70 *tweet*, negatif sebesar 184 *tweet* dan netral sebesar 75 *tweet*.
3. Sentimen masyarakat menggunakan metode komputasional *text mining* dengan algoritma *Decision tree 5.0* pada kata ”Rohingya” di sosial media X menunjukkan bahwa sentimen positif sebanyak 54 *tweet*, negatif sebanyak 140 *tweet* dan netral sebanyak 135 *tweet*.

Hasil pengujian pada 2 metode algoritma *text mining* yaitu algoritma *naïve bayes* dan *decision tree* dengan menggunakan metode *K-Fold Validation* menunjukkan bahwa algoritma *naïve Bayes* mempunyai tingkat akurasi yang lebih baik dengan tingkat akurasi sebesar 83% sedangkan *decision tree* hanya memiliki tingkat akurasi sebesar 78%.

#### UCAPAN TERIMAKASIH

Terima kasih kami berikan kepada Universitas Islam Negeri K.H. Abdurrahman Wahid khususnya dosen dan staff fakultas ekonomi syariah yang telah membantu dalam kelancaran penelitian ini.

#### DAFTAR PUSTAKA

- [1] W. Wasalmi, “Sejarah Konflik Muslim Rohingya,” vol. 4, no. 2, pp. 2964–5476, 2023, doi: 10.59059/tarim.v4i2.217.
- [2] Setwapres, “Wapres Nilai Pengungsi Rohingya sebagai Masalah Kemanusiaan, Harus Diatasi,” <https://www.wapresri.go.id/wapres-nilai-pengungsi-rohingya-sebagai-masalah-kemanusiaan-harus-diatasi-bersama/>, Dec. 05, 2023.
- [3] A. W. Tambunan, “Kerja Sama UNHCR dan IOM dalam Menangani Pencari Suaka dan Pengungsi Etnis Rohingya di Indonesia,” *Journal of International Relations*, vol. 5, no. 2, pp. 341–350, 2019, [Online]. Available: [- s1.undip.ac.id/index.php/jihiWebsite:<http://www.fisip.undip.ac.id>
  - \[4\] E. Albert and L. Maizland, “The Rohingya Crisis,” \*Council on Foreign Relations.\*, Jan. 2020.
  - \[5\] I. Chaturvedi, \*Sentiment Analysis and Opinion Mining: Theoretical and Practical Perspectives\*. Springer, 2021.
  - \[6\] W. , Medhat, A. Hassan, and H. Korashy, “Sentiment Analysis Algorithms and Applications: A Survey,” \*Journal of Computer Science.\*, 2021.
  - \[7\] C. C. Aggarwal, \*Machine Learning for Text\*. Springer, 2018.
  - \[8\] T. Jo, \*Text mining\*, vol. 45. Cham, Switzerland: Springer International Publishing, 2019.
  - \[9\] H. Alhuzali, T. Zhang, and S. Ananiadou, “Emotions and topics expressed on Twitter during the COVID-19 pandemic in the United Kingdom: Comparative geolocation and text mining analysis.,” \*J Med Internet Res\*, vol. 24, no. 10, 2022.
  - \[10\] S. Demirel, E. Kahraman, and U. Gündüz, “A text mining analysis of the change in status of the Hagia Sophia on Twitter: the political discourse and its reflections on the public opinion,” \*Atl J Commun\*, pp. 1–28, 2022.
  - \[11\] T. Siddiqui, “Sarcasm detection from twitter database using text mining algorithms. Turkish Journal of Computer and Mathematics Education,” \*TURCOMAT\*, vol. 12, no. 11, 2021.
  - \[12\] M. Anandarajan, C. Hill, and T. Nolan, \*Text preprocessing. Practical text analytics:Maximizing the value of text data\*. Springer, 2019.
  - \[13\] M. Lamba and M. Madhusudhan, \*Text Mining for Information Professionals: An Uncharted Territory\*. Springer, 2022.
  - \[14\] A. Aziza and S. Rani, “Klasifikasi Sentimen Radikalisme dalam Konten Dakwah Radikal Indonesia melalui Media Sosial Twitter dengan Menggunakan Analisis Sentimen dan Text Mining,” 2023.
  - \[15\] A. Lubis and M. Y. H. Setyawan, “Analisis Sentimen Terhadap Aplikasi Pospay Menggunakan Algoritma Support Vector Machine dan Naive Bayes,” \*Jurnal Teknologi Dan Sistem Informasi Bisnis\*, vol. 6, no. 3, Jul. 2024, doi: 10.47233/jteksis.v6i3.1310.
  - \[16\] M. G. Wibowo and I. Pratama, “Sentiment Analysis of the Digital Population Identity Application Reviews Using the Support Vector Machine Method,” \*Jurnal Teknologi Dan Sistem Informasi Bisnis\*,](http://ejournal-</a></li></ol></div><div data-bbox=)

- vol. 6, no. 4, Oct. 2024, doi:  
10.47233/jteksis.v6i4.1552.
- [17] H. A. Chowdhury, T. A. Nibir, and M. S. Islam, "Sentiment Analysis of Comments on Rohingya Movement with Support Vector Machine," 2018. [Online]. Available:  
<https://www.researchgate.net/publication/324007412>
- [18] N. Rochmawati and S. C. Wibawa, "Opinion Analysis on Rohingya using Twitter Data," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Apr. 2018. doi: 10.1088/1757-899X/336/1/012013.
- [19] D. Toresa *et al.*, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Mengukur Tingkat Kepuasan Mahasiswa Dalam Penggunaan Edlink," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 5, no. 3, Jul. 2023, doi: 10.47233/jteksis.v5i3.855.
- [20] I. R. Hendrawan, E. Utami, and A. D. Hartanto, "Analisis Perbandingan Metode Tf-Idf dan Word2vec pada Klasifikasi Teks Sentimen Masyarakat Terhadap Produk Lokal di Indonesia," *Jurnal Smart Comp*, vol. 11, no. 3, 2022.