

OPINION MINING TERHADAP PENANGANAN ROHINGYA DI PROVINSI ACEH MENGGUNAKAN NAÏVE BAYES PADA MEDIA SOSIAL TWITTER

Yuliawati^{1*}, Nina Sulistyowati², Siska³

^{1,2,3} sistem Informasi, Universitas Singaperbangsa Karawang; Jl. HS. Ronggo Waluyo, Pseurjaya, Telukjambe Timur, Karawang, Jawa Barat, 41361.

Received: 14 Juli 2024

Accepted: 31 Juli 2024

Published: 7 Agustus 2024

Keywords:

Opinion Mining; Sentiment Analysis; Naïve Bayes; KDD (Knowledge Discovery in Database); Rohingya; Twitter

Correspondent Email:

yuliawati.pwk@gmail.com

Abstrak. Asia Tenggara dikenal dengan kekayaan budaya dan etnis yang beragam, namun kerap menghadapi konflik dan intoleransi antarkelompok. Kasus Rohingya menyoroti dampak negatif dari perbedaan ini, sering dibahas di platform Twitter. Untuk mengidentifikasi gambaran tentang penanganan Rohingya di Aceh, penelitian ini menggunakan metode KDD (Knowledge Discovery in Database). Proses ini mencakup seleksi, preprocessing, transformasi, data mining, dan evaluasi. Studi ini menunjukkan bahwa skenario 90:10 dengan pelabelan Inset Lexicon dan SMOTE menghasilkan akurasi model yang meningkat dari 80% menjadi 83%, dengan presisi 88%, recall 76%, dan f-measure 81%. Penggunaan Naïve Bayes untuk menganalisis sentimen dari 2.584 tweet Twitter menunjukkan hasil yang menjanjikan. Rekomendasi untuk penelitian mendatang mencakup pengumpulan data yang lebih luas dan pertimbangan algoritma seperti Random Forest atau Support Vector Machine. Pemerintah perlu mengadopsi kebijakan inklusif untuk memfasilitasi integrasi Rohingya di Aceh, sementara UNHCR harus meningkatkan advokasi dan kerjasama dengan pemerintah serta meningkatkan kesadaran masyarakat tentang integrasi pengungsi.

Abstract. Southeast Asia is known for its rich cultural and ethnic diversity, but often faces conflict and intolerance between groups. The Rohingya case highlights the negative impact of these differences, often discussed on the Twitter platform. To identify descriptions of the handling of the Rohingya in Aceh, this research uses the KDD (Knowledge Discovery in Database) method. This process includes selection, preprocessing, transformation, data mining, and evaluation. This study shows that the 90:10 scenario with Inset Lexicon and SMOTE labeling resulted in improved model accuracy from 80% to 83%, with 88% precision, 76% recall, and 81% f-measure. The use of Naïve Bayes to analyze the sentiment of 2,584 Twitter tweets showed promising results. Recommendations for future research include wider data collection and consideration of algorithms such as Random Forest or Support Vector Machine. The government needs to adopt inclusive policies to facilitate the integration of Rohingya in Aceh, while UNHCR should increase advocacy and cooperation with the government and raise public awareness about refugee integration.

1. PENDAHULUAN

Asia Tenggara dikenal dengan keanekaragaman budaya dan etnis yang melimpah. Variasi agama dan etnis di wilayah ini mencerminkan kemajemukan yang kompleks. Keanekaragaman ini sering kali menjadi sumber konflik dan intoleransi sosial. Mayoritas etnis sering kali mendominasi minoritas, yang kadang-kadang mengarah pada pengungsian kelompok minoritas dari wilayah mereka. Contoh nyata dari dampak keanekaragaman ini adalah masalah pengungsi, seperti kelompok Rohingya yang menjadi perhatian global karena keadaan mereka yang terus memburuk [1].

Pengungsi Rohingya telah menjadi isu kontemporer yang mendesak dan mengkhawatirkan, dengan sejumlah besar pengungsi yang mencari perlindungan di berbagai negara, termasuk Indonesia. Di Aceh, khususnya, kasus-kasus kedatangan pengungsi Rohingya telah menimbulkan pro dan kontra di kalangan masyarakat lokal. Meskipun beberapa penduduk Aceh menentang kedatangan mereka, yang lain menganggap perlunya koordinasi dan tanggapan yang manusiawi terhadap situasi ini [2].

Pemerintah, dalam hal ini UNHCR, berperan penting dalam memberikan perlindungan dan advokasi bagi pengungsi Rohingya di Aceh. Dukungan ini menjadi krusial mengingat kompleksitas tantangan integrasi sosial dan ekonomi yang dihadapi oleh pengungsi [3]. Dengan perkembangan media sosial, platform seperti Twitter telah menjadi kanal penting bagi masyarakat untuk menyuarakan pendapat dan pandangan mereka tentang isu-isu kontroversial seperti pengungsi Rohingya. Oleh karena itu, analisis sentimen dari data media sosial dapat memberikan wawasan yang berharga untuk mendukung kebijakan dan tindakan yang lebih inklusif dan berkelanjutan di masa depan [4].

Algoritma Naive Bayes adalah salah satu metode klasifikasi yang sangat populer dalam analisis data dan teks, sering digunakan untuk mengolah data teks singkat seperti dalam representasi persepsi masyarakat. Metode ini terkenal karena sederhana namun memiliki akurasi dan kinerja yang tinggi dalam pengklasifikasian teks [5]. Di bidang data

mining, confusion matrix menjadi salah satu metode utama untuk mengukur performa model klasifikasi dengan menghitung akurasi, presisi, dan tingkat kesalahan algoritma. Kemajuan dalam teknologi informasi memungkinkan pengumpulan data dalam skala besar dan pengolahan data yang cepat, meningkatkan efisiensi dalam analisis dan pengambilan keputusan berbasis data [6].

Berdasarkan pembahasan sebelumnya maka dilakukan sebuah penelitian dengan judul “Opinion Mining Terhadap Penanganan Rohingya di Provinsi Aceh” dengan menggunakan algoritma *Naïve Bayes*.

2. TINJAUAN PUSTAKA

Analisis sentimen adalah teknik untuk memahami dan mengklasifikasikan emosi (baik, negatif, dan netral) dalam teks tertulis, dengan menggunakan strategi evaluasi pesan. Metode ini juga mencakup pengategorian opini sebagai positif, netral, atau negatif. Sebagai alat otomatis, analisis sentimen digunakan untuk memproses data teks dan menghasilkan informasi yang relevan. Ini memungkinkan identifikasi pendapat individu terhadap subjek dan objek dalam dataset [7].

Twitter adalah platform media sosial yang terkenal di kalangan pengguna internet karena kemudahannya. Selain itu, pengguna dapat secara bebas berbagi pendapat dan informasi di platform ini [8]. Twitter bukan hanya tempat untuk berinteraksi dan berkomunikasi, tetapi juga digunakan untuk menyuarakan aspirasi serta menyediakan informasi tentang peristiwa-peristiwa aktual yang sedang terjadi dalam masyarakat [9].

Data mining adalah proses yang melibatkan serangkaian tahap. Tahap awal melibatkan pengumpulan data dari sumber yang relevan, kemudian diikuti oleh tahap pemrosesan yang bertujuan untuk meningkatkan kualitas dan mengubah data. Tahap akhir dari proses ini melibatkan interpretasi dan evaluasi data yang menghasilkan informasi baru atau pengetahuan [10].

Text mining dan data mining dianggap sebagai bidang yang sama, karena algoritma yang dapat digunakan dalam kedua bidang. Namun, keduanya berbeda karena teks mining memerlukan preprocessing dan melibatkan

elemen tertentu dan tidak terstruktur, sementara data mining menggunakan data terstruktur. Bidang *Natural Language Processing* (NLP) dan *text mining* berhubungan [11].

Klasifikasi *Naïve Bayes* adalah suatu metode klasifikasi yang berdasarkan pada teorema Bayesian dan probabilitas, dengan mengasumsikan bahwa setiap variabel x adalah independen. Dalam data mining, ada beberapa metode yang dapat digunakan untuk mengevaluasi kinerja model yang dihasilkan. Metode confusion matrix biasanya digunakan untuk mengevaluasi performa algoritma atau model klasifikasi dengan menunjukkan keakuratan, presisi, dan tingkat kesalahan algoritma [12].

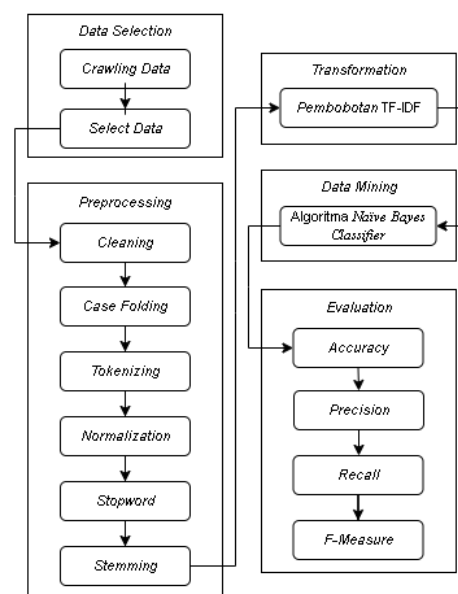
Masyarakat di Asia Tenggara memiliki banyak keanekaragaman budaya dan etnis. Variasi agama dan etnis di Asia Tenggara mencerminkan kemajemukan. Karena kemajemukan ini, kelompok mayoritas etnis dan kelompok minoritas dibagi menjadi kelompok-kelompok di seluruh Negara. Salah satu masalah yang sedang ricuh yang dihadapi masyarakat global saat ini adalah masalah pengungsi orang asing di dalam negeri [1].

Perlu perhatian khusus dalam penanganan isu pengungsi saat ini, termasuk dalam konteks kelompok etnis Rohingya yang terus meningkat jumlahnya ketika mereka mencari perlindungan di berbagai negara. Masalah sosial dan ekonomi dapat muncul sebagai akibat dari banyaknya pengungsi yang tiba di Indonesia. Jika pemerintah tidak mengontrol mereka dengan baik, pengungsi juga dapat meningkatkan pengangguran dan kriminalitas [3].

3. METODE PENELITIAN

Metodologi penelitian yang digunakan adalah KDD (*Knowledge Discovery in Database*). Proses KDD adalah proses yang mengidentifikasi pattern dalam data yang benar, unik, berguna, dan dimengerti. Tahapan pada proses KDD (*Knowledge Discovery in Database*) terdiri dari selection, preprocessing, transformation, data mining, dan interpretation/evaluation [13]. Alur dari metode *Knowledge Discovery in Database* (KDD) dengan menggunakan algoritma *Naïve Bayes*. Pemrosesan dimulai dengan tahap pemilihan data (data selection), yaitu proses *crawling* data dan pemilihan data yang relevan. Tahap berikutnya adalah *preprocessing* yang

mencakup proses *cleaning*, *case folding*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming*. Setelah *preprocessing*, dilakukan transformasi data dengan pembobotan menggunakan metode TF-IDF. Selanjutnya, proses data mining dilakukan menggunakan algoritma *Naïve Bayes Classifier*. Tahap akhir dari penelitian ini adalah evaluasi (*evaluation*) untuk menentukan nilai akurasi (*accuracy*), presisi (*precision*), *recall*, dan *f-measure* dari model yang telah dibangun.



Gambar 1. Metodologi Penelitian

3.1. Data Selection

Dalam penelitian ini, dilakukan pencarian dan ekstraksi data dari platform media sosial Twitter. Proses pengambilan data menggunakan teknik *crawling* data diimplementasikan melalui bahasa pemrograman *Python*, dengan rentang waktu pengambilan data dari tanggal 01 Desember 2023 hingga 30 April 2024. Selanjutnya, data diambil dengan menggunakan kata kunci "pengungsi rohingya", yang akan mencakup semua tweet yang berisi pendapat masyarakat terhadap penanganan Rohingya di Twitter.

3.1.1. Crawling Data

Crawling dilakukan dengan menggunakan *crawler* atau *spider* yang mengikuti tautan antarhalaman, mengekstrak informasi dari struktur web, dan mengumpulkan data berdasarkan kata kunci

atau rentang waktu tertentu. Hasilnya berupa dataset yang berisi teks, gambar, atau informasi lain yang penting untuk mendukung penelitian ini.

3.1.2. Select Data

Seleksi data dari proses hasil *crawling* dilakukan dengan mempertimbangkan kriteria tertentu seperti kata kunci yang relevan atau rentang waktu spesifik.

3.2. Preprocessing

Preprocessing teks adalah tahap di mana aplikasi melakukan seleksi data yang akan diproses pada setiap dokumen. Setelah data tweet telah melewati tahap seleksi, langkah berikutnya adalah tahap *preprocessing* untuk mendapatkan data yang bersih agar dapat diolah pada tahap selanjutnya.

3.2.1. Cleaning

Proses pembersihan (*cleaning*) melibatkan penghapusan URL, angka, dan simbol-simbol yang tidak diperlukan dalam proses klasifikasi, seperti link yang tercantum di tweet dan emotikon – emotikon yang tidak diperlukan. Contoh hasil proses *cleaning* ditunjukkan pada Tabel 1.

Tabel 1. Hasil Cleaning

Sebelum Cleaning	Sesudah Cleaning
@dewilingga19 @CIImtgr Gua juga islam tapi gua gak setuju rohingya nya di Indonesia ini bukan pengungsi tapi perdagangan orang	Gua juga islam tapi gua gak setuju rohingya nya di Indonesia ini bukan pengungsi tapi perdagangan orang
Pengungsi Rohingya di Indonesia semakin banyak Anji teringat Palestina. https://t.co/9gwpphiHYX	Pengungsi Rohingya di Indonesia semakin banyak Anji teringat Palestina httpstco.gwpphiHYX
@dulhomot @laira_salsabila @folkshittmedia Lah bule datang kan sebage turis ya bawa duit mayan pemasukan negaraRohingya pengungsi ilegal alias abal wajar sih kena usir	Lah bule datang kan sebage turis ya bawa duit mayan pemasukan negaraRohingya pengungsi ilegal alias abal wajar sih kena usir

3.2.2. Case Folding

Tidak semua dokumen teks menggunakan huruf kapital dengan cara yang seragam. Maka dari itu, Folding Case memiliki peran penting dalam mengubah seluruh teks dokumen menjadi bentuk standar, umumnya dalam bentuk huruf kecil atau lowercase.

Contoh hasil proses *case folding* ditunjukkan pada Tabel 2.

Tabel 2. Hasil Case Folding

Sebelum Case Folding	Sesudah Case Folding
@dewilingga19 @CIImtgr Gua juga islam tapi gua gak setuju rohingya nya di Indonesia ini bukan pengungsi tapi perdagangan orang	gua juga islam tapi gua gak setuju rohingya nya di indonesia ini bukan pengungsi tapi perdagangan orang
Pengungsi Rohingya di Indonesia semakin banyak Anji teringat Palestina. https://t.co/9gwpphiHYX	pengungsi rohingya di indonesia semakin banyak anji teringat palestina httpstco.gwpphihyx
@dulhomot @laira_salsabila @folkshittmedia Lah bule datang kan sebage turis ya bawa duit mayan pemasukan negaraRohingya pengungsi ilegal alias abal wajar sih kena usir	lah bule datang kan sebage turis ya bawa duit mayan pemasukan negararohingya pengungsi ilegal alias abal wajar sih kena usir

3.2.3. Tokenizing

Tokenisasi umumnya melibatkan pembagian kumpulan karakter dalam teks menjadi unit kata, memungkinkan untuk membedakan karakter tertentu yang berfungsi sebagai pemisah kata atau tidak. Contoh hasil proses *tokenizing* ditunjukkan pada Tabel 3.

Tabel 3. Hasil Tokenizing

Sebelum Tokenizing	Sesudah Tokenizing
@dewilingga19 @CIImtgr Gua juga islam tapi gua gak setuju rohingya nya di Indonesia ini bukan pengungsi tapi perdagangan orang	gua,juga,islam,tapi,gua,gak,setuju ,rohingya,nya,di,indonesia,ini,bu kan,pengungsi,tapi,perdagangan, orang
Pengungsi Rohingya di Indonesia semakin banyak Anji teringat Palestina. https://t.co/9gwpphiHYX	pengungsi,rohingya,di,indonesia, semakin,banyak,anji,teringat,pale stina, httpstco.gwpphihyx
@dulhomot @laira_salsabila @folkshittmedia Lah bule datang kan sebage turis ya bawa duit mayan pemasukan negaraRohingya pengungsi ilegal alias abal wajar sih kena usir	lah,bule,dateng,kan,sebage,turis,y a,bawa,duit,mayan,pemasukan,ne gararohingya,pengungsi,ilegal,ali as,abal,wajar,sih,kena,usir

3.2.4. Normalization

Tokenisasi umumnya melibatkan pembagian kumpulan karakter dalam teks menjadi unit kata, memungkinkan untuk membedakan karakter tertentu yang berfungsi sebagai pemisah kata atau tidak. Contoh hasil proses *tokenizing* ditunjukkan pada Tabel 4.

Tabel 4. Hasil Normalization

Sebelum Normalization	Sesudah Normalization
@dewilingga19 @CImtgr Gua juga islam tapi gua gak setuju rohingya nya di Indonesia ini bukan pengungsi tapi perdagangan orang	gua,juga,islam,tapi,gua,gak,setuju ,rohingya,nya,di,indonesia,ini,bu kan,pengungsi,tapi,perdagangan, orang
Pengungsi Rohingya di Indonesia semakin banyak Anji teringat Palestina. https://t.co/9gwpphiHYX	pengungsi,rohingya,di,indonesia, semakin,banyak,anji,teringat,pale stina,httpstco,gwpphihyx
@dulhomot @laira_salsabila @folkshittmedia Lah bule dateng kan sebage turis ya bawa duit mayan pemasukan negara.Rohingya pengungsi ilegal alias abal wajar sih kena usir	lah,bule,dateng,kan,sebage,turis,y a,bawa,duit,mayan,pemasukan,ne gararohingya,pengungsi,ilegal,ali as,abal,wajar,sih,kena,usir

3.2.5. Stopword

Proses filtrasi melibatkan ekstraksi kata-kata penting dari hasil token. Metode ini dapat menggunakan algoritma untuk stoplist (mengeluarkan kata-kata yang tidak penting) atau wordlist (mempertahankan kata-kata yang relevan). Contoh *stopwords* termasuk kata-kata seperti "yang," "dan," "di," "dari," dan sebagainya. Contoh hasil proses *stopword* ditunjukkan pada Tabel 5.

Tabel 5. Hasil Stopword

Sebelum Stopword	Sesudah Stopword
@dewilingga19 @CImtgr Gua juga islam tapi gua gak setuju rohingya nya di Indonesia ini bukan pengungsi tapi perdagangan orang	islam,setuju,rohingya,nya,indone sia,pengungsi,perdagangan,orang
Pengungsi Rohingya di Indonesia semakin banyak Anji teringat Palestina. https://t.co/9gwpphiHYX	pengungsi,rohingya,indonesia,anj i,palestina,httpstco,gwpphihyx
@dulhomot @laira_salsabila @folkshittmedia Lah bule dateng kan sebage turis ya bawa duit mayan pemasukan negara.Rohingya pengungsi ilegal alias abal wajar sih kena usir	bule,turis,ya,bawa,duit,lumayan,p emasukan,negararohingya,pengu ngsi,ilegal,alias,palsu,wajar,sih,k ena,usir

3.2.6. Stopword

Proses mengubah kata-kata dalam dokumen menjadi bentuk kata dasar melibatkan penghapusan semua imbuhan, termasuk awalan (prefix), akhiran (suffix), sisipan (infix), dan gabungan awalan-akhiran (confix). Contoh

hasil proses *stemming* ditunjukkan pada Tabel 6.

Tabel 6. Hasil Stemming

Sebelum Stemming	Sesudah Stemming
@dewilingga19 @CImtgr Gua juga islam tapi gua gak setuju rohingya nya di Indonesia ini bukan pengungsi tapi perdagangan orang	islam,tuju,rohingya,nya,indonesia ,ungsi,dagang,orang
Pengungsi Rohingya di Indonesia semakin banyak Anji teringat Palestina. https://t.co/9gwpphiHYX	ungsi,rohingya,indonesia,anji,pal estina,httpstco,gwpphihyx
@dulhomot @laira_salsabila @folkshittmedia Lah bule dateng kan sebage turis ya bawa duit mayan pemasukan negara.Rohingya pengungsi ilegal alias abal wajar sih kena usir	bule,turis,ya,bawa,duit,lumayan,p asu,negararohingya,ungsi,ilegal,a lias,palsu,wajar,sih,kena,usir

3.3. Transformation

Dalam tahapan transformasi, terjadi seleksi fitur dengan mengubah nilai dari kategori menjadi numerik. Proses transformasi ini menggunakan metode TF-IDF, di mana pemilihan fitur mempertimbangkan jumlah kata yang muncul dalam dokumen.

3.3.1. Pembobotan TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) adalah teknik pembobotan yang menentukan pentingnya kata dalam dokumen relatif terhadap kumpulan dokumen lainnya. TF mengukur frekuensi kata dalam dokumen, sedangkan IDF mengukur kepentingan kata di seluruh dokumen, dengan kata yang sering muncul di banyak dokumen memiliki IDF rendah. Dengan mengalikan TF dan IDF, maka akan didapatkan bobot TF-IDF, yang menunjukkan bahwa kata yang sering muncul dalam satu dokumen tetapi jarang di dokumen lain memiliki bobot tinggi.

3.4. Transformation

Pada tahap ini, data tweet diproses dengan mengkategorikannya dalam dua kategori: sentimen positif dan negatif. Ini dilakukan dengan menggunakan algoritma pengklasifikasi *Naïve Bayes*. Data sebelumnya telah dibagi menjadi dua bagian: data pelatihan dan data pengujian. Tiga skenario berbeda digunakan

untuk data pelatihan dan pengujian pada saat ini, yaitu 90:10, 80:20, 70:30.

3.4.1. Algoritma Naïve Bayes

Pemrosesan algoritma Naive Bayes melibatkan beberapa tahapan utama. Pertama, dilakukan pengumpulan dan pra-pemrosesan dataset, termasuk tokenisasi, penghapusan stop words, dan stemming. Selanjutnya, dataset dibagi menjadi data pelatihan dan data pengujian. Prediksi dibuat menggunakan teorema Bayes untuk menghitung probabilitas posterior, sehingga data dapat diklasifikasikan ke dalam kelas dengan probabilitas tertinggi. Terakhir, performa model dievaluasi dengan metrik seperti akurasi, presisi, recall, dan F-measure.

3.5. Evaluation

Evaluasi adalah tahap terakhir dalam proses Knowledge Discovery in Database (KDD), di mana hasil pengolahan data dievaluasi untuk memastikan pola informasi dapat dipahami dengan baik. Tahap ini penting dalam KDD karena menilai prediksi mesin dan akurasi data asli.

3.5.1. Accuracy

Perhitungan akurasi pada tahap evaluasi melibatkan membandingkan prediksi model dengan data aktual untuk menilai seberapa tepat model tersebut. Akurasi dihitung dengan membagi jumlah prediksi yang benar (baik positif maupun negatif) dengan total jumlah prediksi yang dibuat.

3.5.2. Precision

Perhitungan presisi pada tahap evaluasi merupakan cara untuk menilai keakuratan model dalam mengidentifikasi instance positif dari semua prediksi yang dihasilkan. Presisi dihitung dengan membagi jumlah instance positif yang benar diprediksi oleh model dengan total prediksi positif yang dilakukan.

3.5.3. Recall

Perhitungan recall pada tahap evaluasi mengukur kemampuan model untuk mengidentifikasi semua instance positif dalam dataset. Ini dihitung dengan membagi jumlah true positive (prediksi yang benar untuk kelas positif) dengan jumlah true positive ditambah

false negative (instance positif yang tidak terdeteksi).

3.5.4. F-Measure

Perhitungan F-measure adalah metrik yang mengkombinasikan presisi dan recall menjadi satu nilai tunggal untuk mengevaluasi efektivitas suatu model klasifikasi. F-measure sangat berguna ketika terjadi trade-off antara keakuratan prediksi positif (presisi) dan kemampuan model untuk menemukan semua instance positif (recall). Formula F-measure diperoleh dari perhitungan harmonik antara presisi dan recall, memberikan gambaran holistik tentang performa model dalam mengklasifikasikan data.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan algoritma *Naïve Bayes*. Metodologi yang diterapkan adalah *Knowledge Discovery in Database* (KDD), yang mencakup *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation*. Teknik pengumpulan data yang digunakan adalah crawling data menggunakan Google Colaboratory. Pada rentang waktu 1 Desember 2023 hingga 30 April 2024, berhasil dikumpulkan total 2.584 tweet. Setelah tahap penyeleksian, terpilih 1.915 tweet yang terdiri dari 978 tweet positif dan 937 tweet negatif.

Setelah itu, data diubah dengan menggunakan pembobotan kata menggunakan Term Frequency (TF). Langkah berikutnya melibatkan proses data mining dengan algoritma *Naïve Bayes*, yang dievaluasi dalam tiga skenario berbeda: 90:10, 80:20, dan 70:30. Evaluasi akhir dilakukan dengan menggunakan confusion matrix untuk mengevaluasi kinerja algoritma yang telah digunakan.

Hasil Perbandingan *confusion matrix* sebelum dan sesudah menggunakan penerapan SMOTE ditunjukkan pada Tabel 4.10. di bawah ini.

Tabel 7. Perbandingan Hasil Confusion Matrix

Pelatihan	Pembagian Data	Penerapan SMOTE	Confusion Matrix			
			Accuracy	Precision	Recall	F-Measure
Manual	90:10	Tidak	79%	78%	76%	77%
		Ya	74%	73%	78%	75%
	80:20	Tidak	77%	78%	74%	76%
		Ya	73%	71%	76%	73%
	70:30	Tidak	73%	73%	71%	72%
		Ya	72%	71%	74%	72%
Inset Lesscon	90:10	Tidak	80%	73%	41%	53%
		Ya	83%	88%	76%	81%
	80:20	Tidak	80%	73%	41%	52%
		Ya	82%	86%	76%	80%
	70:30	Tidak	79%	78%	40%	52%
		Ya	81%	86%	74%	79%

Dalam penelitian ini, hasil akurasi yang terbaik dicapai pada skenario pembagian data 90:10 dengan penggunaan pelabelan Inset Lexicon setelah dilakukan SMOTE. Sebelum proses SMOTE, akurasi model mencapai 80%, yang mengalami peningkatan menjadi 83% setelah penerapan SMOTE, menunjukkan peningkatan sebesar 3% dengan precision sebesar 88%, recall sebesar 76%, dan f-measure sebesar 81%. Hasil ini menggambarkan bahwa penggunaan SMOTE efektif dalam meningkatkan kinerja model dalam mengatasi ketidakseimbangan data, terutama ketika digunakan bersama metode pelabelan yang lebih kompleks seperti *Inset Lexicon*.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan beberapa hal, diantaranya:

- a. Penelitian ini berhasil melakukan analisis sentimen masyarakat terhadap penanganan Rohingya di platform Twitter menggunakan algoritma Naïve Bayes. Temuan dari analisis menunjukkan bahwa pendekatan ini efektif dalam mengidentifikasi pola opini positif dan negatif terkait dengan respons terhadap penanganan pengungsi di Provinsi Aceh. Pendekatan ini memberikan pemahaman yang mendalam mengenai pandangan dan sikap publik terhadap isu yang sedang banyak dibicarakan di era digital.
- b. Pengambilan opini tentang penanganan Rohingya di Provinsi Aceh dilakukan dengan crawling menggunakan Python di Google Colaboratory. Data dari tweet dengan kata kunci "pengungsi Rohingya" dari 1 Desember 2023 hingga 30 April 2024 terkumpul sebanyak 2.584 tweet. Data kemudian melewati proses seleksi dari data awal sebanyak 2.584 menjadi sebanyak 1.915 data. Selanjutnya, data tersebut dilabeli menjadi sentimen positif dan negatif menggunakan pelabelan manual oleh ahli bahasa serta otomatis dengan Inset Lexicon. Hasilnya, pelabelan manual didapatkan 937 sentimen positif dan 978 sentimen negatif, sedangkan Inset Lexicon menunjukkan 454 sentimen positif dan 1.265 sentimen negatif. Hasil analisis yang didapatkan didominasi sentimen negatif, menunjukkan bahwa

kedatangan pengungsi Rohingya di Aceh kurang disambut oleh masyarakat.

- c. Hasil evaluasi menunjukkan bahwa algoritma Naïve Bayes diuji dalam tiga skenario pembagian data: 90:10, 80:20, dan 70:30. Evaluasi menggunakan confusion matrix untuk mengukur kinerja algoritma menunjukkan hasil terbaik pada skenario 90:10 setelah penerapan SMOTE dan pelabelan Inset Lexicon. Sebelum SMOTE, akurasi model adalah 80%, meningkat menjadi 83% setelah penerapan SMOTE, dengan precision 88%, recall 76%, dan f-measure 81%. Ini menunjukkan bahwa SMOTE efektif dalam meningkatkan kinerja model untuk mengatasi ketidakseimbangan data, terutama ketika digunakan bersama metode pelabelan yang lebih kompleks seperti Inset Lexicon.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberikan dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] S. Nisrina and R. Wachid, "Peran International Organization for Mitigation (IOM) dalam Menangani," *Jurnal Independen*, p. 41, 2023.
- [2] D. Nada, "Peran Pemerintah Dalam Menangani Pengungsi Rohingya Di Kota Lhokseumawe (Studi pada Kantor Imigrasi Kelas II TPI kota Lhokseumawe)," *Kajian Administrasi Negara: Riset Dan Pengabdian*, p. 46, 2021.
- [3] A. Syahrin, "Dimensi Hak Asasi Manusia Dalam Penanganan Kasus Pengungsi Rohingya: Pendekatan Hukum Interdisipliner," *Lex Librum : Jurnal Ilmu Hukum*, pp. 898-899, 2019.
- [4] D. Darwis, E. Pratiwi and A. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Jurnal Ilmiah Educat*, 2020.
- [5] D. Darwis, N. Siskawati and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *Jurnal TEKNO KOMPAK*, vol. 15, pp. 131-145, 2021.

- [6] C. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 3, pp. 145-160, 2022.
- [7] I. Rahman, A. Hasanah and N. Heryana, "Analisis Sentimen Ulasan Pengguna Aplikasi Samsat Digital Nasional (Signal) Dengan Menggunakan Metode Naïve Bayes Classifier," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, pp. 963-969, 2024.
- [8] M. Azahri, N. Sulistiyowati and M. Jajuli, "Analisis Sentimen Pengguna Kereta Api Indonesia Melalui Sosial Media Twitter dengan Algoritma Naïve Bayes Classifier," *JATI (Jurnal Mahasiswa Teknik Informatika)*, 2023.
- [9] C. Fadilah and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, 2022.
- [10] U. S, "Penerapan Data Mining Dengan Mengimplementasikan Algoritma K-Means Dalam Proses Clustering Untuk Pengelompokan Mahasiswa Calon Penerima Beasiswa KIP," *Building of Informatics, Technology and Science (BITS)*, pp. 70-85, 2023.
- [11] A. Fauziyyah and D. Gautama, "Analisis Sentimen Pandemi Covid-19 Pada Streaming Twitter Dengan Text Mining Python," *Jurnal Ilmiah Sinus (JIS)*, 2020.
- [12] A. Fairuz, A. Ramadhani and N. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," *Jurnal DINDA (Indonesian Journal of Data Science, IOT, Machine Learning and Artificial Intelligence)*, 2021.
- [13] Nurhachita and E. Negara, "A Comparison Between Naïve Bayes and The K-Means Clustering Algorithm for The Application of Data Mining on The Admission of New Students," *Jurnal Intelektualita: Keislaman, Sosial, dan Sains*, 2020.