



Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes

Dhea Ananda, Ryan Randy Suryono*

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

Email: ¹dhea_ananda@teknokrat.ac.id, ^{2,*}ryan@teknokrat.ac.id

Email Penulis Korespondensi: ryan@teknokrat.ac.id

Abstrak—Kedatangan pengungsi Rohingya di Indonesia telah menjadi topik yang penuh kontroversi dan memicu berbagai respons dari masyarakat. Dalam konteks ini, analisis sentimen publik terhadap kedatangan pengungsi Rohingya menjadi sangat penting untuk memahami dinamika perasaan, opini, dan sikap masyarakat Indonesia terhadap isu ini. Dalam menjalankan analisis sentimen publik, pemilihan metode menjadi krusial untuk memastikan hasil yang akurat. Tujuan dari penelitian ini untuk melakukan analisis sentimen publik terkait kedatangan pengungsi Rohingya dengan menggunakan metode support vector machine (SVM) dan naïve bayes. Fokus utamanya adalah mengevaluasi sentimen publik serta membandingkan kinerja kedua metode tersebut. Dua metode yang umum digunakan dalam analisis sentimen adalah Support Vector Machine (SVM) dan Naïve Bayes. Penelitian ini menggunakan dataset sebanyak 3350 tweet untuk melakukan analisis sentimen publik terhadap kedatangan pengungsi Rohingya di Indonesia. Dalam penelitian ini, dilakukan pembagian data menggunakan metode pembagian 70:30, di mana 70% dari data digunakan untuk pelatihan model dan 30% digunakan untuk pengujian model. Hasil penelitian menunjukkan bahwa model SVM memiliki akurasi sebesar 76%, sedangkan model Naïve Bayes memiliki akurasi sebesar 70%. Hasil penelitian menunjukkan bahwa model SVM memiliki akurasi sebesar 76%, sementara model Naïve Bayes memiliki akurasi sebesar 70%. Hal ini menunjukkan bahwa model SVM mampu memprediksi sentimen dengan lebih baik dan memiliki angka kesalahan yang lebih rendah dibandingkan dengan model Naïve Bayes.

Kata Kunci: Analisis Sentimen; Twitter; SVM; Naïve Bayes; Rohingya

Abstract—The arrival of Rohingya refugees in Indonesia has become a highly controversial topic, eliciting various responses from the public. In this context, public sentiment analysis regarding the arrival of Rohingya refugees is crucial for understanding the dynamics of feelings, opinions, and attitudes of the Indonesian society towards this issue. In conducting public sentiment analysis, the selection of methods is crucial to ensure accurate results. The aim of this research is to conduct sentiment analysis regarding the arrival of Rohingya refugees using the Support Vector Machine (SVM) and Naïve Bayes methods. The main focus is to evaluate public sentiment and compare the performance of both methods. Two common methods used in sentiment analysis are Support Vector Machine (SVM) and Naïve Bayes. This research utilized a dataset of 3350 tweets to conduct public sentiment analysis on the arrival of Rohingya refugees in Indonesia. In this study, data was divided using the 70:30 split method, where 70% of the data was used for model training and 30% for model testing. The research findings indicate that the SVM model has an accuracy of 76%, while the Naïve Bayes model has an accuracy of 70%. This suggests that the SVM model is better at predicting sentiments and has lower error rates compared to the Naïve Bayes model.

Keywords: Sentiment Analysis; Twitter; SVM; Naïve Bayes; Rohingya

1. PENDAHULUAN

Rohingya adalah kelompok etnis minoritas muslim yang masih dianggap imigran ilegal, sebagian besar tinggal di bagian Rakhine di Myanmar [1]. Rohingya telah terlibat dalam konflik yang berkepanjangan, yang melibatkan diskriminasi, pembatasan hak, serta tindakan kekerasan dari pihak militer. Sebagai akibatnya, mereka telah melarikan diri secara massal ke beberapa negara di sekitar wilayah Bangladesh, Thailand, Malaysia, dan Indonesia [2]. Indonesia sebagai negara kepulauan yang besar juga terlibat dalam menanggapi krisis pengungsi ini. Kedatangan pengungsi Rohingya di Indonesia menjadi topik yang penuh kontroversi dan mengundang berbagai respons dari masyarakat. Dalam hal ini, analisis sentimen publik terhadap kedatangan pengungsi Rohingya menjadi sangat penting untuk memahami dinamika perasaan, opini, dan sikap masyarakat Indonesia terkait isu ini [3]. Beberapa kelompok masyarakat mungkin bersikap simpatik dan bersedia membantu, sedangkan kelompok lain merasa khawatir akan dampak sosial dan ekonomi yang mungkin ditimbulkan. Media sosial terutama twitter, memainkan peran penting dalam analisis sentimen terkait isu-isu sosial dan kemanusiaan [4]. Melalui jumlah besar komentar dan kemudahan akses, twitter memberikan wawasan mendalam terhadap perasaan dan pandangan masyarakat Indonesia terhadap masalah seperti kedatangan pengungsi Rohingya. Dengan ragam komentar, penggunaan hashtag, dan bahasa informal, analisis sentimen di twitter memberikan pemahaman secara menyeluruh dan real-time terhadap komentar masyarakat [5].

Dalam menjalankan analisis sentimen publik, pemilihan metode menjadi penting untuk memastikan hasil yang akurat. Dua metode yang umum digunakan dalam analisis sentimen adalah support vector machine (SVM) dan naïve bayes, penggunaan dua metode ini memungkinkan untuk melakukan perbandingan langsung antara kinerja keduanya, sehingga memungkinkan penelitian ini untuk mengevaluasi keunggulan dan kelemahan dari masing-masing pendekatan. [6]. Penelitian yang memanfaatkan model Support Vector Machine (SVM) dan Naïve Bayes pada dataset sebanyak 5000 tweet menunjukkan bahwa SVM dengan kernel RBF mencapai akurasi tertinggi, mencapai 88,8%, dibandingkan dengan kernel Linear dan Polynomial. Sementara itu, model Naïve Bayes mencapai akurasi sebesar 82,51% [7]. Berdasarkan penelitian yang membahas analisis sentiment tentang dampak pandemi virus corona di twitter, menggunakan metode klasifikasi sentimen naïve bayes dan support vector



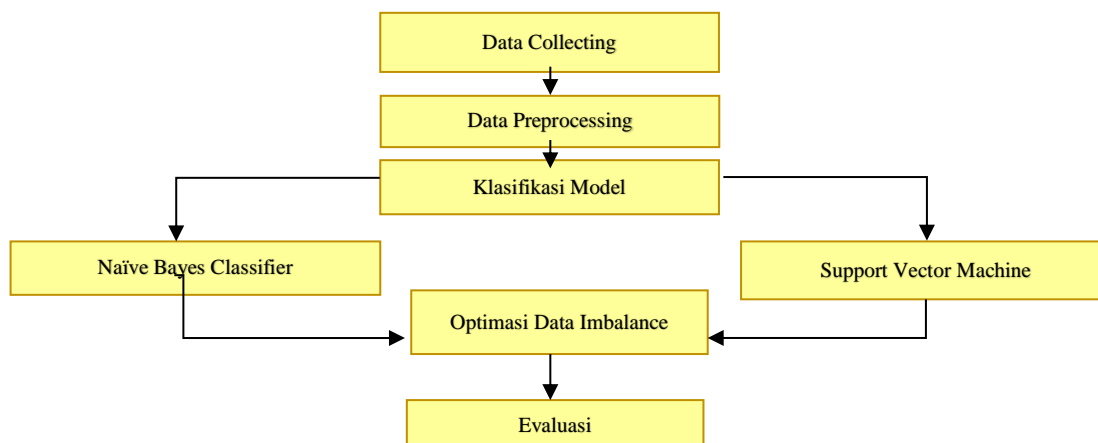
machine. Dataset terdiri dari 1104 tweet dengan tiga sentiment yaitu positif, negatif, dan netral. Hasil pengujian menunjukkan akurasi Support Vector Machine sebesar 79,96%, sedangkan Naïve Bayes mencapai 81,07%. Dengan demikian, naïve bayes lebih unggul dalam klasifikasi sentimen terkait virus corona di Twitter [6]. Selanjutnya mengenai penelitian yang membahas tentang BTS sebuah grup vokal yang berasal dari Korea Selatan dan sedang populer di kalangan remaja Indonesia. Penelitian ini membandingkan dua pendekatan penelitian, yaitu Naïve Bayes dan Support Vector Machine, untuk mengevaluasi akurasi. Hasil penelitian menunjukkan bahwa akurasi menggunakan metode Naïve Bayes mencapai 79%, sedangkan metode Support Vector Machine mencapai 81%. Dari kedua metode tersebut, Support Vector Machine memiliki tingkat akurasi yang lebih tinggi. Analisis sentimen menunjukkan bahwa mayoritas komentar yang diperoleh dari pengguna Twitter bersifat positif [8].

Dalam penelitian analisis sentimen terhadap ChatGPT yang dilakukan dengan menggunakan dataset berjumlah 1000 sampel, hasilnya menunjukkan bahwa model SVM dengan label RoBERTa mencapai akurasi sebesar 55%, sedangkan Naïve Bayes mencapai akurasi sebesar 47% [9]. Analisis sentimen dengan melakukan scraping ulasan pengguna aplikasi Shopee, yang berhasil mengumpulkan 3000 data ulasan. Penelitian menunjukkan bahwa penerapan algoritma Support Vector Machine menghasilkan kinerja yang sangat baik, dengan tingkat akurasi mencapai 98% dan nilai F1-score sebesar 0,98, yang juga setara dengan 98% [10]. Penelitian menggunakan dua metode klasifikasi yang berbeda, yaitu Naïve Bayes dan SVM. Kedua metode ini digunakan untuk memprediksi label pada dataset berdasarkan pola data yang ada di dalamnya. Hasil uji coba menunjukkan bahwa Naïve Bayes memiliki kinerja yang lebih baik dengan tingkat akurasi mencapai 91,48%, sementara SVM mencapai tingkat akurasi sebesar 85,47% [11]. SVM adalah algoritma klasifikasi yang kuat yang dapat menangani data dalam berbagai dimensi dan mampu menemukan batas keputusan yang optimal untuk memisahkan kelas sentimen yang berbeda. Naïve Bayes merupakan metode yang sederhana namun efektif, yang didasarkan pada teorema Bayes dengan asumsi independensi antara fitur-fitur dalam data [12]. Meskipun sederhana, Naïve Bayes dapat memberikan hasil yang baik dalam klasifikasi sentimen dengan akurasi yang tinggi. Dengan memilih metode yang sesuai dengan karakteristik data dan kebutuhan spesifik penelitian, analisis sentimen dapat dilakukan dengan lebih efektif dan menghasilkan hasil yang dapat diandalkan [13].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan analisis sentimen publik terkait kedatangan pengungsi Rohingya menggunakan metode Support Vector Machine (SVM) dan Naïve Bayes [14]. Fokus utamanya adalah mengevaluasi sentimen publik terhadap kedatangan pengungsi Rohingya. Manfaat dari penelitian ini termasuk memberikan pemahaman yang lebih baik tentang kinerja dan perbandingan antara dua metode analisis sentimen yang umum digunakan, yaitu SVM dan Naïve Bayes. Hasil yang akan dicapai dalam penelitian ini mencakup evaluasi kinerja kedua metode analisis sentimen terkait kedatangan pengungsi Rohingya di Indonesia, serta perbandingan akurasi, presisi, recall, dan F1-Score dari kedua metode tersebut. Dengan demikian, penelitian ini diharapkan dapat memberikan wawasan yang lebih baik tentang efektivitas metode analisis sentimen dan membantu pengambilan keputusan yang lebih baik dalam menangani isu-isu sosial yang kompleks. Melalui penelitian ini, diharapkan dapat memberikan kontribusi pada pemahaman yang lebih baik tentang persepsi masyarakat terhadap pengungsi Rohingya.

2. METODOLOGI PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah pendekatan kuantitatif, di mana sampel diambil dari platform Twitter dengan menggunakan kata kunci "Rohingya". Tahapan penelitian ini mencakup pengumpulan data, preprocessing, klasifikasi model menggunakan metode Naïve Bayes dan Support Vector Machine, dilanjutkan dengan penerapan teknik optimasi data imbalance untuk mengatasi ketidakseimbangan data, serta evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Berikut ini adalah tahapan penelitian yang dilakukan oleh penulis dalam penelitian ini.



Gambar 1. Tahapan Penelitian



2.1 Data Collection

Data collection adalah proses pengumpulan informasi atau data dari berbagai sumber yang relevan dengan tujuan penelitian atau analisis yang sedang dilakukan. Pengambilan data dari Twitter melibatkan proses ekstraksi informasi mengenai pengguna dan tweet dari platform tersebut, dengan menggunakan kata kunci spesifik [15]. Ekstraksi data ini dilakukan melalui pemanfaatan Application Programming Interface (API). Dataset pada penelitian ini diperoleh melalui penerapan teknik scrapping dengan menggunakan kata kunci "Rohingya". Pengumpulan data dilakukan dengan memanfaatkan teknik scrapping data, menggunakan library tweet harvest yang tersedia dalam bahasa pemrograman python. Data yang terkumpul sebanyak 3350 data. Dataset tersebut kemudian dilakukan preprocessing, yakni membersihkan data, menghapus karakter yang tidak perlu, pelabelan data.

2.2 Preprocessing

Preprocessing adalah proses persiapan data yang dilakukan sebelum data tersebut diolah lebih lanjut atau dianalisis [16]. Tahapan ini sangat penting karena bertujuan untuk membersihkan dataset yang masih dalam keadaan tidak terstruktur. Proses ini melibatkan beberapa tahapan seperti pembersihan data, case folding, tokenization, stopwords, dan stemming. Tujuan dari proses ini adalah menghasilkan dataset yang bersih, terstruktur, dan siap untuk analisis lebih lanjut. Dengan melakukan langkah-langkah tersebut, dataset menjadi lebih mudah dipahami dan diproses oleh model analisis, sehingga memungkinkan untuk mendapatkan wawasan yang lebih akurat dan informatif dari data yang ada.

2.3 Cleaning Data

Cleaning bertujuan untuk menghilangkan karakter atau simbol seperti tautan URL, nama pengguna atau mention (@username), hashtag (#), retweet, dan emotikon dari teks. Setelah proses cleaning dilakukan, kalimat menjadi lebih bersih tanpa adanya emotikon, sehingga memudahkan dalam proses analisis data dan pemrosesan lanjutan dapat dilihat pada tabel 1.

Tabel 1. Cleaning Data

Tweet	Cleaning Data
Kebencian terhadap pengungsi Rohingya ini banyak terjadi karena misinformasi dan propaganda. Ada beberapa kebohongan, misinformasi, dan berita aneh yang aku temui dan bisa diverifikasi kebenarannya. Jika ada yang menyerukan kebencian terhadap Rohingya, bisa pake rujukan ini 📺📺 Apakah benar pengungsi Rohingya tidak tahu terima kasih? Tidak juga. Berkali-kali mengucapkan terima kasih karena bantuan orang Indonesia. Anggapan tidak berterima kasih ini perlu jelas, karena tuduhannya kepada seluruh pengungsi. https://t.co/JenbhCZLHM https://t.co/K1un8MLdOw https://t.co/jCw2kusFBD	Kebencian terhadap pengungsi Rohingya ini banyak terjadi karena Ada beberapa berita aneh aku temui bisa Jika ada kebencian terhadap Rohingya bisa pake rujukan ini Apakah benar pengungsi Rohingya tidak tahu terima kasih Tidak juga terima kasih karena bantuan orang Indonesia Anggapan tidak berterima kasih ini perlu jelas karena kepada seluruh pengungsi

2.4 Case Folding

Case folding adalah proses mengubah semua teks dalam dokumen menjadi huruf kecil karena tidak semua teks konsisten dalam penggunaan huruf kapital. Proses ini memperhatikan sensitivitas terhadap huruf besar dan kecil. Berikut adalah hasil tahapan case folding yang dapat dilihat pada tabel 2.

Tabel 2. Case Folding

Twitter	Case Folding
Kebencian terhadap pengungsi Rohingya ini banyak terjadi karena Ada beberapa berita aneh aku temui bisa Jika ada kebencian terhadap Rohingya bisa pake rujukan ini Apakah benar pengungsi Rohingya tidak tahu terima kasih Tidak juga terima kasih karena bantuan orang Indonesia Anggapan tidak berterima kasih ini perlu jelas karena kepada seluruh pengungsi	kebencian terhadap pengungsi rohingya ini banyak terjadi karena ada beberapa berita aneh aku temui bisa jika ada kebencian terhadap rohingya bisa pake rujukan ini Apakah benar pengungsi Rohingya tidak tahu terima kasih Tidak juga terima kasih karena bantuan orang Indonesia Anggapan tidak berterima kasih ini perlu jelas karena kepada seluruh pengungsi

2.5 Tokenizing Data

Setelah proses pembersihan, tahapan selanjutnya adalah tokenizing. Pada tahap ini, data akan diproses dengan menghilangkan tanda baca sehingga menghasilkan kumpulan kata atau kalimat yang berdiri sendiri. Berikut adalah hasil tahapan tokenizing yang dapat dilihat pada tabel 3.



Tabel 3. Tokenzing Data

Tweet	Tokenization
kebencian terhadap pengungsi rohingya ini banyak terjadi karena ada beberapa berita aneh aku temui bisa jika ada kebencian terhadap rohingya bisa pake rujukan ini	['kebencian', 'terhadap', 'pengungsi', 'rohingya', 'ini', 'banyak', 'terjadi', 'karena', 'ada', 'beberapa', 'berita', 'aneh', 'aku', 'temui', 'bisa', 'jika', 'ada', 'kebencian', 'terhadap', 'rohingya', 'bisa', 'pake', 'rujukan', 'ini']
Apakah benar pengungsi Rohingya tidak tahu terima kasih Tidak juga terima kasih karena bantuan orang Indonesia Anggapan tidak berterima kasih ini perlu jelas karena kepada seluruh pengungsi	['apakah', 'benar', 'pengungsi', 'rohingya', 'tidak', 'tahu', 'terima', 'kasih', 'tidak', 'juga', 'terima', 'kasih', 'karena', 'bantuan', 'orang', 'indonesia', 'anggapan', 'tidak', 'berterima', 'kasih', 'ini', 'perlu', 'jelas', 'karena', 'kepada', 'seluruh', 'pengungsi']

2.6 Stop Word

Tahapan stopword bertujuan untuk menghilangkan kata-kata umum yang sering muncul dalam jumlah besar dan tidak memiliki makna penting. Hal ini dapat dilakukan dengan menggunakan algoritma daftar kata bantu untuk menghapus kata-kata yang kurang relevan, atau menggunakan daftar kata penting untuk menyimpan kata-kata yang signifikan. Berikut adalah hasil tahapan stopwords yang dapat dilihat pada tabel 4.

Tabel 4. Stop Word

Tweet	Stop Word
['kebencian', 'terhadap', 'pengungsi', 'rohingya', 'ini', 'banyak', 'terjadi', 'karena', 'ada', 'beberapa', 'berita', 'aneh', 'aku', 'temui', 'bisa', 'jika', 'ada', 'kebencian', 'terhadap', 'rohingya', 'bisa', 'pake', 'rujukan', 'ini']	['kebencian', 'terhadap', 'pengungsi', 'rohingya', 'banyak', 'terjadi', 'karena', 'ada', 'beberapa', 'berita', 'aneh', 'temui', 'bisa', 'jika', 'ada', 'kebencian', 'terhadap', 'rohingya', 'bisa', 'pake', 'rujukan']
['apakah', 'benar', 'pengungsi', 'rohingya', 'tidak', 'tahu', 'terima', 'kasih', 'tidak', 'juga', 'terima', 'kasih', 'karena', 'bantuan', 'orang', 'indonesia', 'anggapan', 'tidak', 'berterima', 'kasih', 'ini', 'perlu', 'jelas', 'karena', 'kepada', 'seluruh', 'pengungsi']	['apakah', 'benar', 'pengungsi', 'rohingya', 'tidak', 'tahu', 'terima', 'kasih', 'tidak', 'terima', 'kasih', 'karena', 'bantuan', 'orang', 'indonesia', 'anggapan', 'tidak', 'berterima', 'kasih', 'perlu', 'jelas', 'karena', 'kepada', 'seluruh', 'pengungsi']

2.7 Stemming

Dalam proses stemming, setiap kata akan diubah menjadi bentuk dasarnya dengan menghilangkan imbuhan. Stemming yang digunakan untuk melakukan stemming adalah library sastrawi. Berikut adalah hasil tahapan stemming yang dapat dilihat pada tabel 5.

Tabel 5. Stemming

Tweet	Stemming Data
['kebencian', 'terhadap', 'pengungsi', 'rohingya', 'banyak', 'terjadi', 'karena', 'ada', 'beberapa', 'berita', 'aneh', 'temui', 'bisa', 'jika', 'ada', 'kebencian', 'terhadap', 'rohingya', 'bisa', 'pake', 'rujukan']	benci hadap ungsi rohingya banyak jadi karena ada beberapa berita aneh temu bisa jika ada benci hadap rohingya bisa pake rujuk
['apakah', 'benar', 'pengungsi', 'rohingya', 'tidak', 'tahu', 'terima', 'kasih', 'tidak', 'terima', 'kasih', 'karena', 'bantuan', 'orang', 'indonesia', 'anggapan', 'tidak', 'berterima', 'kasih', 'perlu', 'jelas', 'karena', 'kepada', 'seluruh', 'pengungsi']	apakah benar ungsi rohingya tidak tahu terima kasih tidak terima kasih karena bantu orang indonesia anggap tidak terima kasih perlu jelas karena kepada seluruh ungsi

2.8 Pelabelan Data

Pada tahap ini, dataset akan diberi label positif, dan negatif. Penentuan label pada dataset dilakukan secara manual dengan memberi sentiment mengenai komentar yang ada di twitter. Berikut adalah hasil tahapan pelabelan data yang dapat dilihat pada tabel 6.

Tabel 6. Pelabelan Data

Tweet	Pelabelan Data
rohingya berita hidup damai indonesia gimana cara supaya sbg rakyat dengar amp dr rohingya segera keluar dr tanah indonesia kayak gerak lambat bahkan kes mau ulur agar ttp biar stay sini sumpah ngeri bgt makin lama makin bnyk jumlah	Positif
	Negatif



2.9 Future Extraction (TF – IDF)

Future Extraction atau Ekstraksi fitur diartikan sebagai salah satu teknik yang sering digunakan untuk mengurangi kompleksitas data di mana data yang besar diubah menjadi fitur – fitur yang lebih sederhana namun tetap memiliki nilai informasi. Penambahan fitur ini dapat secara signifikan meningkatkan kinerja model. Term Frequency (TF) dan Inverse Document Frequency (IDF) digunakan untuk menentukan rasio berapa kali suatu kata muncul dalam suatu dokumen dibandingkan dengan jumlah kata secara keseluruhan dalam dokumen tersebut.

$$TF(t, d) = \frac{\text{Jumlah kemunculan term } t \text{ dalam dokumen } d}{\text{Total jumlah term dalam dokumen } d} \quad (1)$$

TF mengukur seberapa sering suatu term muncul dalam suatu dokumen.

2.10 SMOTE

Teknik SMOTE (Synthetic Minority Over-sampling Technique) merupakan salah satu strategi untuk menangani ketidakseimbangan kelas dalam dataset dengan menciptakan data sintetis untuk kelas yang kurang representatif [16]. Jumlah data sintetis yang dihasilkan oleh SMOTE mungkin tidak selalu sama dengan jumlah awal data minoritas. Proses ini membantu meningkatkan jumlah sampel dalam kelas minoritas, menyetarakan pengelompokan kelas, dan meminimalkan dampak ketidakseimbangan pada kinerja model klasifikasi [17]. Penggunaan SMOTE menjadi relevan dalam berbagai konteks analisis data, khususnya ketika kelas minoritas memiliki representasi yang terbatas dan dibutuhkan pendekatan oversampling untuk meningkatkan akurasi dan ketepatan model [18].

2.11 Klasifikasi Model

Klasifikasi model adalah representasi matematis dari cara sebuah sistem memisahkan atau mengklasifikasikan data ke dalam kategori atau kelas yang berbeda berdasarkan fitur-fitur yang diberikan. Algoritma Naive Bayes dan Support Vector Machine adalah dua pendekatan yang umum digunakan. Keduanya sering digunakan dalam machine learning karena efisiensi dan keandalannya [19]. Tujuan utama dari model klasifikasi adalah untuk menghasilkan prediksi yang akurat dan konsisten untuk data baru.

2.12 Support Vector Machine

Support Vector Machine (SVM) merupakan suatu metode klasifikasi yang berfungsi memisahkan data ke dalam beragam kelas dengan mencari hiperplane optimal yang memaksimalkan jarak antara kelas-kelas tersebut. Algoritma SVM efektif untuk menyelesaikan permasalahan klasifikasi biner (dua kelas) atau multikelas [20]. Tujuannya adalah untuk menemukan batas keputusan yang optimal yang memisahkan data sebaik mungkin. SVM dapat bekerja dengan baik dalam ruang fitur berdimensi tinggi dan sangat efektif dalam menangani data yang kompleks. Berikut gambar ilustrasi SVM.

2.13 Naïve Bayes Classifier

Naive Bayes classifier adalah algoritma klasifikasi yang didasarkan pada teorema Bayes dengan asumsi independensi antara fitur-fitur yang diberikan. Ini adalah pendekatan yang cepat dan sederhana untuk klasifikasi yang sering digunakan dalam pemrosesan teks, klasifikasi email spam, dan sistem rekomendasi [15]. Naive Bayes memprediksi kelas atau label untuk data baru dengan menghitung probabilitas kelas berdasarkan distribusi fitur-fitur yang diamati. Meskipun sederhana, Naive Bayes dapat memberikan hasil yang baik dalam banyak situasi dan sering digunakan sebagai baseline untuk membandingkan kinerja model klasifikasi yang lebih kompleks.

$$P(X | Y) = (P(Y | X) \cdot P(X)) / (P(Y)) \quad (2)$$

Rumus diatas merupakan landasan dari algoritma klasifikasi Naive Bayes. Dalam konteks ini, $P(X | Y)$ adalah probabilitas kejadian X terjadi jika kejadian Y telah diamati. Ini dihitung dengan mengalikan probabilitas kejadian Y terjadi jika kejadian X telah diamati ($P(Y | X)$) dengan probabilitas prior kejadian X ($P(X)$), dan kemudian dibagi dengan probabilitas prior kejadian Y ($P(Y)$). Probabilitas kondisional, $P(Y | X)$, menunjukkan seberapa mungkin kejadian Y terjadi jika X telah diamati, sementara $P(X)$ adalah seberapa sering X terjadi secara umum. $P(Y)$ mewakili seberapa sering kejadian Y terjadi tanpa mempertimbangkan X. Dengan menggunakan rumus ini, Naive Bayes dapat memperkirakan probabilitas kelas atau label berdasarkan fitur-fitur yang diamati, meskipun dengan asumsi bahwa fitur-fitur tersebut adalah independen satu sama lain. Meskipun sederhana, metode ini sering digunakan dalam klasifikasi teks dan pengenalan pola dengan dataset besar.

2.14 Optimasi Data Imbalance

Optimasi Data Imbalance adalah proses penyesuaian atau penanganan ketidakseimbangan dalam distribusi kelas atau label pada dataset. Ini terjadi ketika satu kelas atau label memiliki jumlah contoh yang jauh lebih banyak atau lebih sedikit dibandingkan dengan kelas atau label lainnya dalam dataset [18]. Optimasi Data Imbalance bertujuan untuk meningkatkan kualitas model pembelajaran mesin dengan mengurangi efek dari ketidakseimbangan tersebut. Metode SMOTE (Synthetic Minority Over-sampling Technique) digunakan dalam optimasi untuk



menyelaraskan dataset dengan meningkatkan jumlah sampel dalam kelas minoritas, dalam hal ini, sentimen negatif, sehingga jumlahnya menjadi sebanding dengan kelas mayoritas, yaitu sentimen positif. Pada awalnya, sentimen negatif hanya terdiri dari 1617 data, sementara sentimen positif memiliki 1733 data. Dengan menerapkan SMOTE, dilakukan pembuatan sampel sintetis baru untuk sentimen negatif. Teknik ini menciptakan sampel baru dengan cara memilih titik-titik data dalam kelas minoritas dan membuat sampel sintetis baru berdasarkan tetangga terdekat. Dengan demikian, jumlah sampel dalam kelas minoritas ditingkatkan hingga mencapai jumlah yang setara dengan kelas mayoritas, yaitu 1733. Dengan mengimbangi jumlah data antara kedua kelas, proses pembelajaran mesin dapat menjadi lebih seimbang dan model dapat memberikan prediksi yang lebih akurat, karena tidak ada bias yang diakibatkan oleh ketidakseimbangan kelas.

2.15 Evaluasi

Langkah berikutnya adalah model diuji melalui penerapan teknik K-fold cross-validation, di mana performa model dinilai berdasarkan metrik kesalahan untuk memperoleh tingkat akurasi. Penilaian kinerja model melibatkan penerapan matriks kebingungan untuk mendapatkan ukuran-ukuran seperti akurasi, presisi, recall, dan f1-Score. Confusion matrix digunakan untuk menganalisis kinerja model pada setiap iterasi. Terdapat empat elemen utama dalam confusion matrix, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) [21]. Komponen-komponen ini digunakan untuk menghitung nilai-nilai dalam confusion matrix yang mencerminkan performa model dalam mengklasifikasikan data. Berdasarkan dari matriks kebingungan tersebut, dapat dihitung nilai akurasi, presisi, recall, dan f1-score dengan persamaan berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (4)$$

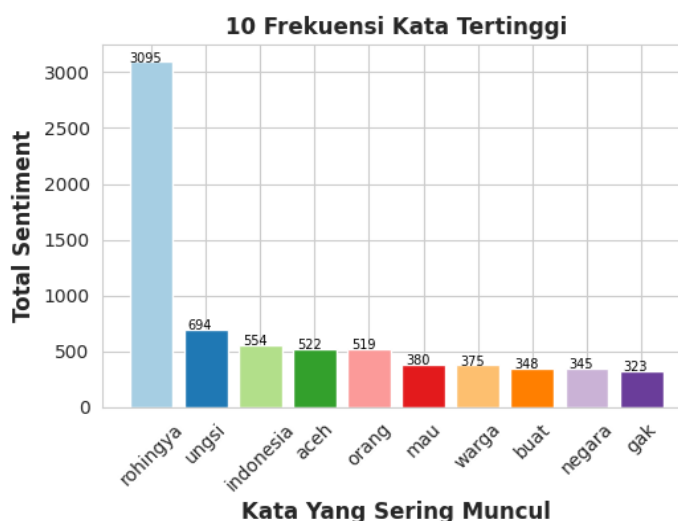
$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

$$f1 - \text{score} = \frac{2 \times \text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Dataset diperoleh dari pengguna pada media platform twitter yang digunakan adalah 3350 tweet, setelah itu dilakukan tahap preprocessing teks pada dataset, pelabelan, dan klasifikasi model. Dalam data tersebut, terdapat 10 kata dengan frekuensi tertinggi dari sentiment berlabel positif dan negatif, antara lain "rohingya", "ungsi", "indonesia", "aceh", "orang", "mau", "warga", "buat", "negara", dan "gak". Untuk visualisasi, digunakan wordcloud untuk semua sentimen, serta wordcloud khusus untuk sentimen positif dan negatif. Berikut frekuensi kata yang sering muncul pada gambar 2.



Gambar 2. Frekuensi Kata

Dalam penelitian ini memperoleh pembagian sentiment yaitu 1733 data sentimen positif dan 1617 data sentimen negatif sehingga terdeteksi ketidakseimbangan kelas, untuk mengatasinya diterapkan metode SMOTE untuk menangani perbedaan jumlah antar kelas. Perbandingan antara data sebelum penerapan SMOTE dapat dilihat pada gambar 3.



Jumlah Klasifikasi Sentimen Sesudah Optimasi SMOTE

Sentiment	Jumlah
Positif	1733
Negatif	1733

3.2 Tahap Visualisasi Word Cloud

[illegible]

Gambar 5. WordCloud Semua Sentiment



3.3 Tahap Pengujian

Pada tahap ini, data yang telah melalui langkah-langkah pra-pemrosesan seperti pembersihan dan normalisasi akan diselidiki lebih lanjut menggunakan dua teknik analisis yang berbeda, yaitu Support Vector Machine (SVM) dan Naïve Bayes. SVM memisahkan data ke dalam kelas-kelas berbeda dengan mencari hiperplane terbaik yang memaksimalkan jarak antara kelas-kelas, sementara Naïve Bayes memanfaatkan teorema Bayes dengan asumsi independensi antara fitur-fitur dalam data untuk melakukan klasifikasi. Pembagian data menggunakan perbandingan 70% data training dan 30% data testing. Dalam tahap pengujian yang akan dibandingkan yaitu nilai accuracy, F1-score, precision, recall, dan confusion matrix. Setelah penerapan SMOTE, terlihat adanya peningkatan pada performa model SVM dalam hal akurasi yang meningkat menjadi 70%, namun terjadi penurunan pada recall menjadi 55%. Di sisi lain, model Naïve Bayes menunjukkan peningkatan yang lebih signifikan, dengan akurasi mencapai 76% dan recall sebesar 73%. Hasil ini menunjukkan bahwa penggunaan SMOTE efektif dalam meningkatkan kinerja model Naïve Bayes dalam mengatasi ketidakseimbangan kelas, sementara pengaruhnya terhadap performa model SVM relatif lebih terbatas. Berikut adalah hasil confusion matrix sebelum SMOTE yang dapat dilihat pada tabel 7, dan sesudah SMOTE pada tabel 8.

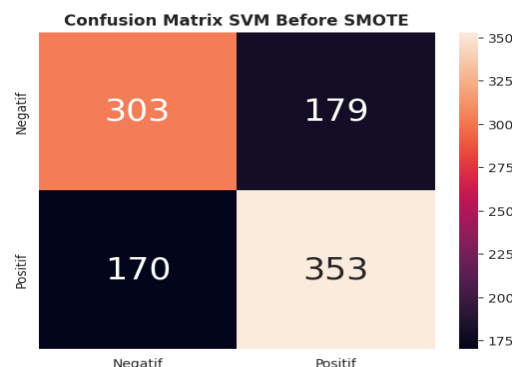
Tabel 7. Hasil Confusion Matrix Sebelum SMOTE

Model	Accuracy	Precision	Recall	f1-score
SVM	65%	64%	63%	63%
Naïve Bayes	65%	61%	78%	68%

Tabel 8. Hasil Confusion Matrix Sesudah SMOTE

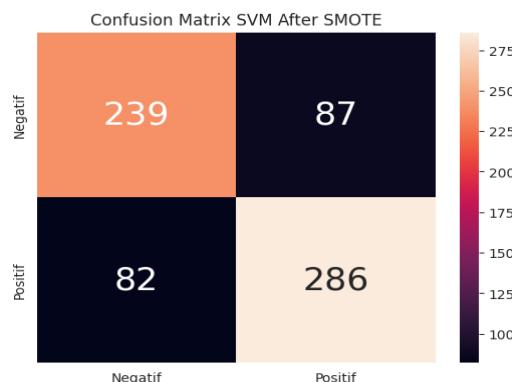
Model	Accuracy	Precision	Recall	f1-score
SVM	70%	75%	55%	63%
Naïve Bayes	76%	74%	73%	74%

Confusion matrix merupakan tabel yang digunakan untuk menilai performa dari suatu model klasifikasi. Tabel ini memperbandingkan prediksi yang dibuat oleh model dengan nilai aktual dalam data pengujian. Dalam penelitian ini, digunakan matriks 2x2 untuk evaluasi. Berikut hasil dari confusion matrix.



Gambar 6. Hasil Confusion Matrix Algoritma Support Vector Machine sebelum optimasi SMOTE

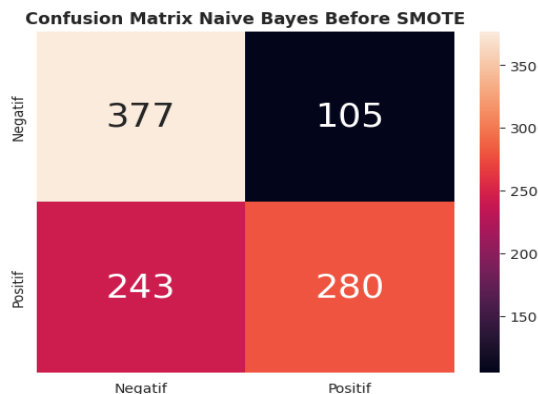
Pada gambar 6 menunjukkan confusion matrix tersebut, terdapat 303 prediksi negatif yang benar dan 179 prediksi positif yang salah untuk kelas negatif. Di sisi lain, terdapat 170 prediksi negatif yang salah dan 353 prediksi positif yang benar untuk kelas positif. Ini menunjukkan kinerja model dalam mengklasifikasikan instance ke dalam kelas positif dan negatif, dengan beberapa kesalahan dalam prediksi kelas positif dan negatif.



Gambar 7. Hasil Confusion Matrix Algoritma Support Vector Machine sesudah optimasi SMOTE

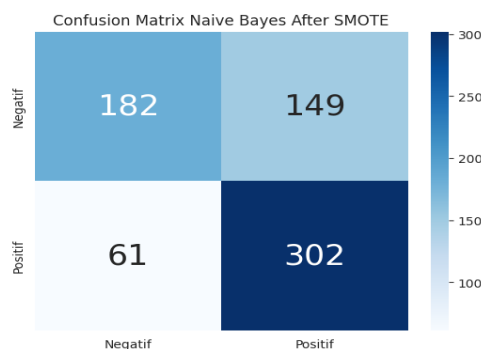


Hasil Confusion matrix pada gambar 7 sebelum menggunakan SMOTE pada algoritma SVM menunjukkan kecenderungan model kurang akurat dalam mengenali sentiment positif terkait rohingya. Dengan 239 True Positives dan 87 False Negatives, model lebih sukses mengidentifikasi sentiment negative (286 True Negatives). Hasil ini menunjukkan bahwa model perlu melakukan peningkatan kemampuan model dalam mengenali sentiment positif.



Gambar 8. Hasil Confusion Matrix Algoritma Naïve Bayes sebelum optimasi SMOTE

Pada gambar 8 confusion matrix tersebut, terdapat 280 prediksi positif yang benar (True Positives) dan 243 prediksi negatif yang salah (False Negatives) untuk kelas positif, sementara terdapat 105 prediksi negatif yang salah (False Positives) dan 377 prediksi negatif yang benar (True Negatives) untuk kelas negatif. Hal ini memberikan gambaran tentang kinerja model dalam mengklasifikasikan instance ke dalam kelas positif dan negatif.



Gambar 9. Hasil Confusion Matrix Algoritma Naïve Bayes sesudah optimasi SMOTE

Pada gambar 9 model memiliki tingkat akurasi yang tinggi dalam mengidentifikasi sentimen positif dengan 302 prediksi positif yang benar (True Positives), namun memiliki tingkat keakuratan yang lebih rendah dalam mengenali sentimen negatif, dengan 182 prediksi negatif yang benar (True Negatives). Setelah menerapkan SMOTE pada algoritma SVM, terjadi peningkatan dalam kemampuan model untuk mengenali sentimen positif terkait Rohingya. Namun, meskipun demikian, model masih cenderung lebih akurat dalam mengidentifikasi sentimen negatif daripada positif. Dari 377 kasus negatif, 105 salah terklasifikasikan sebagai positif, sementara dari 280 kasus positif, 243 salah terklasifikasikan sebagai negatif.

4. KESIMPULAN

Berdasarkan penelitian diatas optimasi dengan metode SMOTE meningkatkan performa model dalam memprediksi sentimen pada dataset yang digunakan. Sebelum optimasi, model Naive Bayes dan SVM memiliki akurasi yang relatif rendah, dengan cenderung melakukan kesalahan dalam memprediksi sentimen tertentu. Setelah dilakukan optimasi dengan SMOTE, terjadi peningkatan signifikan dalam nilai akurasi, precision, recall, dan f1-score pada kedua model. Meskipun terjadi peningkatan performa pada kedua model setelah optimasi SMOTE, model SVM menunjukkan hasil yang lebih baik dibandingkan dengan model Naive Bayes. Dengan nilai akurasi mencapai 76%, precision 0.74, recall 0.73, dan f1-score 0.74, model SVM mampu memprediksi sentimen dengan lebih baik dan memiliki angka kesalahan yang lebih rendah dibandingkan dengan model Naive Bayes.

REFERENCES

- [1] A. Inayah and G. Karisma, "Respon Asean Terhadap Permasalahan Etnis Rohingya Di Myanmar," J. Hub. Int. Indones.,



- vol. 4, no. 1, pp. 31–44, 2022.
- [2] B. Budaya, “Dampak Kewarganegaraan Etnis Rohingya Di Myanmar Terhadap Pelanggaran Hak Asasi Manusia Dan Negara Sekitar,” *J. Ilm. Huk.*, vol. 11, no. 1, pp. 106–120, 2018, [Online]. Available: <https://maksigama.wisnuwardhana.ac.id/index.php/maksigama/article/view/44>
- [3] D. G. S. Mangku, “Pemenuhan Hak Asasi Manusia Kepada Etnis Rohingya Di Myanmar,” *Perspekt. Huk.*, vol. 21, no. 1, pp. 1–15, 2021.
- [4] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, T. Tukino, and B. Priyatna, “Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter,” *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 8, no. 1, p. 11, 2023, doi: 10.31328/jointecs.v8i1.4265.
- [5] R. Maria, R. U. Umayah, S. Mahardinny, D. N. Kalana, and D. D. Saputra, “Analisis Sentimen Persepsi Masyarakat Terhadap Penggunaan Aplikasi My Pertamina Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier,” *J. Komput. Antart.*, vol. 1, no. 1, pp. 1–10, 2023, [Online]. Available: <https://ejournal.mediaantartika.id/index.php/jka/article/view/1%0Ahttps://ejournal.mediaantartika.id/index.php/jka/article/download/1/1>
- [6] C. F. Hasri and D. Alita, “Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022.
- [7] S. Styawati, A. R. Isnain, N. Hendrastuty, and L. Andraini, “Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 1, pp. 56–60, 2021, doi: 10.30591/jpit.v6i1.3245.
- [8] Rina Noviana and Isram Rasal, “Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter,” *J. Tek. dan Sci.*, vol. 2, no. 2, pp. 51–60, 2023, doi: 10.56127/jts.v2i2.791.
- [9] D. Atmajaya, A. Febrianti, and H. Darwis, “Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter,” *Indones. J. Comput. Sci.*, vol. 12, Aug. 2023, doi: 10.33022/ijcs.v12i4.3341.
- [10] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [11] T. T. Widowati and M. Sadikin, “Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.
- [12] Y. Ansori and K. F. H. Holle, “Perbandingan Metode Machine Learning dalam Analisis Sentimen Twitter,” *J. Sist. dan Teknol. Inf.*, vol. 10, no. 4, p. 429, 2022, doi: 10.26418/justin.v10i4.51784.
- [13] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, “Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter,” *J. Inf. Syst. Manag.*, vol. 3, no. 2, pp. 44–49, 2021, doi: 10.24076/joism.2021v3i2.558.
- [14] P. K. Sari and R. R. Suryono, “Komparasi Algoritma Support Vector Machine , Naïve,” *J. Mnemon.*, vol. 2, no. 2, pp. 7–13, 2024.
- [15] K. S. Putri, I. R. Setiawan, and A. Pambudi, “Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier,” *Technol. J. Ilm.*, vol. 14, no. 3, p. 227, 2023, doi: 10.31602/tji.v14i3.11259.
- [16] T. A. Dewi and E. Mailoa, “Perbandingan Implementasi Metode Smote Pada Algoritma Support Vector Machine (Svm) Dalam Analisis Sentimen Opini Masyarakat Tentang Mixue,” *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 3, pp. 849–855, 2023, doi: 10.35870/jimik.v4i3.289.
- [17] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, “Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.
- [18] N. S. Ramadhanti, W. A. Kusuma, and A. Annisa, “Optimasi Data Tidak Seimbang pada Interaksi Drug Target dengan Sampling dan Ensemble Support Vector Machine,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, p. 1221, 2020, doi: 10.25126/jtiik.2020762857.
- [19] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, “Comparison of the Naïve Bayes Method and Support Vector Machine on Twitter Sentiment Analysis,” *SMATIKA J. STIKI Inform. J.*, vol. 10, no. 2, pp. 71–76, 2020.
- [20] M. H. Wicaksono, M. D. Purbolaksono, and S. Al Faraby, “Perbandingan Algoritma Machine Learning untuk Analisis Sentimen Berbasis Aspek pada Review Female Daily,” *eProceedings Eng.*, vol. 10, no. 3, pp. 3591–3600, 2023, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/20631/19944%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/20631>
- [21] A. F. Setyaningsih, D. Septiyani, and S. R. Widiarsari, “Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia,” *J. Teknol. Inform. dan Komput.*, vol. 9, no. 1, pp. 224–235, 2023, doi: 10.37012/jtik.v9i1.1409.