

Analisis Sentimen Pengguna YouTube Tentang Rohingya Menggunakan Algoritma SVM (Support Vector Machine)

Hidayat^{1✉}, Firman Santoso², Lukman Fakhri Lidimillah³

^{1,2,3} Teknologi Informasi, Sains Dan Teknologi, Universitas Ibrahimy Sukorejo Situbondo, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 03-06-2024

Direvisi : 09-06-2024

Diterima : 13-06-2024

Kata Kunci:

Analisis Sentimen, SVM,
YouTube, Rohingya

Keywords :

Sentiment Analysis, SVM,
YouTube, Rohingya.

ABSTRAK

YouTube merupakan platform yang banyak digunakan masyarakat Indonesia, tentu karena terdapat konten-konten yang selalu diperbarui mengikuti berita terkini di Indonesia, salah satunya kasus pengungsi dari Rohingya. Kasus pengungsi Rohingya menimbulkan perbincangan yang luas di platform YouTube di Indonesia. Dalam penelitian ini, kami menerapkan *text mining* untuk menganalisis komentar-komentar yang muncul terkait topik tersebut. Metode SVM dipilih untuk analisis sentimen karena kepopulerannya dan tingkat akurasi yang tinggi. Langkah-langkah meliputi pengumpulan data, pra-pemrosesan data, dan penerapan SVM untuk klasifikasi. Hasilnya menunjukkan akurasi sebesar 77%, dengan *recall* 100% dan *precision* 77%. Temuan ini memberikan wawasan tentang sentimen masyarakat terhadap kasus Rohingya di YouTube. Metode SVM terbukti efektif dalam mengorganisir komentar-komentar yang tidak terstruktur. Penelitian ini dapat menjadi landasan untuk memahami opini dan sikap masyarakat terkait isu ini secara lebih mendalam.

ABSTRACT

YouTube is a platform that is widely used by Indonesian people, of course because there is content that is always updated following the latest news in Indonesia, one of which is the case of Rohingya refugees. The case of Rohingya refugees has caused widespread discussion on the YouTube platform in Indonesia. In this research, we apply text mining to analyze comments that appear regarding this topic. The SVM method was chosen for sentiment analysis because of its popularity and high level of accuracy. Steps include data collection, data pre-processing, and application of SVM for classification. The results show an accuracy of 77%, with recall of 100% and a precision of 77%. These findings provide insight into public sentiment towards the Rohingya case on YouTube. SVM method proved effective in organizing unstructured comments. This research can be a basis for understanding people's opinions and attitudes regarding this issue in more depth.

Corresponding Author :

Hidayat

Prodi Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy Sukorejo Situbondo, Indonesia

Jl. KHR. Syamsul Arifin No. 1-2, Sukorejo

Email: hidayatar78@gmail.com

PENDAHULUAN

Perkembangan teknologi telah mengubah cara kita berinteraksi dan mengakses informasi. Salah satu dampak utamanya adalah kemudahan dalam pertukaran informasi antar perangkat, mengatasi hambatan jarak dan waktu. Platform media sosial seperti Facebook, Twitter, dan YouTube menjadi sarana utama bagi berbagai informasi, termasuk berita nasional dan

internasional. Teknologi juga membantu proses pemindahan informasi dari perangkat satu dengan perangkat yang lain. (Widianto, 2021) Sehingga dengan adanya teknologi ini dunia tidak lagi kesulitan karena batas jarak, maupun waktu untuk mendapatkan informasi (Izzah Nur Masyithoh & Ivo Novitaningtyas, 2021). Beberapa media internet seperti Facebook, Twitter, YouTube dan yang lain-lain merupakan sebuah platform yang menyediakan sebuah informasi baik nasional ataupun internasional. YouTube merupakan sebuah platform yang menyediakan informasi berupa video, sehingga pengguna juga dapat membagikan video-video berbagai jenis seperti video musik, edukasi, tutorial dan lain-lainnya. YouTube juga sangat berperan penting dalam penyebaran konten berita yang sedang terjadi, karena pengguna YouTube lebih mudah dalam menyampaikan sebuah berita dan penggunaanya juga sangat luas. (Di et al., 2024)

Berita yang mengakibatkan terjadinya kehebohan di Masyarakat Indonesia salah satunya adalah kedatangan pengungsi Rohingya di daerah Aceh. Berbagai *channel* YouTube juga banyak yang membahas tentang kedatangan pengungsi Rohingya, dikabarkan dalam berita tersebut bahwasanya pengungsi Rohingya ini melakukan pengungsian secara ilegal, yang mana pengungsi Rohingya ini berdatangan dengan menggunakan kapal yang terbuat dari kayu yang berlabuh di pantai-pantai sekitar provinsi Aceh. Banyak netizen yang berkomentar dalam kejadian itu, baik berkomentar secara positif ataupun *negative*, yang nantinya akan dilakukan penelitian dengan analisis sentimen.

Teks *mining* merupakan salah satu cabang ilmu komputer, teks *mining* memiliki pembahasan tentang analisis sentimen, kegunaan *analysis* sentimen ini adalah untuk memproses data teks menjadi data yang dapat dipahami, (Luqyana et al., 2018) seperti untuk memahami respons positif atau negative dalam data teks yang diproses. (Saputra, 2024) dalam penelitian kali ini algoritma SVM merupakan algoritma yang sangat cocok untuk digunakan, sebab algoritma SVM memiliki akurasi yang tinggi. Seperti penelitian yang telah dilakukan menggunakan perbandingan algoritma SVM dan algoritma naïve bayes, algoritma SVM memiliki akurasi 99,50% sedangkan algoritma naïve bayes akurasi yang dihasilkan 99,25%, (Maulana & Fahmi, 2024) dan juga penelitian yang menggunakan perbandingan algoritma SVM dan *Decision Tree* yang menghasilkan akurasi algoritma SVM sebesar 93,7% sedangkan algoritma *Decision Tree* menghasilkan akurasi sebesar 83%. (Zelina & Afiyati, 2024).

Dari beberapa berita pengungsi Rohingya yang tersebar penulis mengumpulkan data komentar dari beberapa video yang di ambil dari 5 *channel* berita di YouTube, yakni dari *channel* CNN Indonesia, tvOneNews, metro tv, Kompas tv, dan liputan 6. Dari masing-masing *channel* tersebut tentunya banyak komentar netizen yang berbeda baik itu berkomentar positif ataupun berkomentar negatif. Supaya data teks komentar video YouTube dari beberapa *channel* tersebut menjadi informasi yang mudah dipahami, penulis akan melakukan penelitian yang berjudul ANALISIS SENTIMEN PENGGUNA YOUTUBE TENTANG ROHINGYA MENGGUNAKAN ALGORITMA SVM (SUPPORT VECTOR MACHINE). Dengan demikian penelitian ini akan memberikan wawasan yang lebih dalam tentang respon masyarakat terhadap isu ini di platform YouTube

METODE PENELITIAN

Jenis Penelitian

Metode penelitian kuantitatif merupakan metode yang cocok dalam penelitian ini. Metode kuantitatif dapat digunakan sebagai metode analisis yang dapat memahami persepsi pengguna YouTube. Pada penelitian ini algoritma SVM digunakan sebagai pengklasifikasian untuk mengukur serta menganalisis terhadap pengguna YouTube tentang berita pengungsi Rohingya. Teknik statistik dan komputasi digunakan pada penelitian ini guna mengelompokkan komentar pengguna YouTube ke dalam dua kategori yakni sentimen positif dan sentimen negatif. Dengan menerapkan algoritma SVM sebagai analisis dan mengukur sentimen ini bertujuan untuk mengukur dan mengevaluasi dalam menafsirkan sentimen pengguna YouTube terhadap berita

pengungsi rohingya, sehingga dengan penelitian ini dapat menambah wawasan terhadap pengguna YouTube yang lebih mendalam terhadap berita pengungsi rohingya yang terjadi.

Pengumpulan Data

Data yang digunakan berjumlah 2000 data komentar, yang diambil dari 5 *channel* YouTube yang membahas berita tentang pengungsi Rohingya, yakni *channel* tvOneNews, KOMPASTV, Liputan6, METRO TV, dan CNN Indonesia. Proses pengumpulan data dilakukan menggunakan platform Google Colab untuk *crawling* data. Setelah melakukan *crawling* data, kami menjalankan beberapa proses sebelum sampai pada tahap pelabelan. Pada proses pelabelan, kami mengkategorikan data tersebut ke dalam sentimen positif dan sentimen negatif.

Studi Pustaka

Studi pustaka ini bertujuan untuk memfasilitasi pemahaman mendalam terhadap pengetahuan dan penelitian yang relevan, serta membantu peneliti menentukan metode yang paling tepat untuk diterapkan. Pengumpulan data dilakukan melalui analisis dokumen yang konsisten dengan topik penelitian, termasuk informasi dari penelitian terdahulu yang berkaitan dengan tema yang sama. Proses ini memungkinkan peneliti untuk menggali berbagai sumber informasi yang relevan dan meninjau literatur yang telah ada sebelumnya.

Text mining

Text mining termasuk salah satu cabang dari data *mining* yang bertujuan untuk menggali atau mengekstrak dari sejumlah besar teks untuk menjadikan informasi yang bermanfaat. (Kamilla et al., 2024) Didalam proses ini dilakukan eksplorasi dan analisis data teks yang besar dan tidak terstruktur dengan memanfaatkan perangkat lunak untuk mengungkap konsep, pola, topik, kata kunci, dan atribut lainnya dalam data teks tersebut. (Muis & Muhammad, 2023) Namun tujuan terpenting dalam *text mining* adalah bertujuan untuk mendukung dalam menemukan proses untuk menambah pengetahuan pada dokumen yang besar. (Andzar, 2023)

Analisis Sentimen

Analisis sentimen adalah penelitian komputasional yang sering digunakan peneliti terhadap opini, sentimen, dan emosi yang terungkap dalam teks. Analisis sentimen juga merupakan instrumen yang diperlukan untuk mengkategorikan sebuah informasi kedalam sentimen positif ataupun sentimen negatif. (Sentimen et al., 2023) Melakukan analisis sentimen menggunakan pendekatan *text mining* memang merupakan langkah yang tepat untuk memahami dan mengenali makna dari setiap teks opini yang diperoleh. Pendekatan ini memanfaatkan teknik-teknik dari data *mining* dan pemrosesan bahasa alami untuk mengekstrak informasi penting dari teks, seperti sentimen atau perasaan yang terkandung di dalamnya. (Sutresno, 2023)

SVM (Support Vector Machine)

Metode ini diperkenalkan oleh Vapnik sebagai algoritma yang efisien untuk klasifikasi terhadap permasalahan nonlinear pada tahun 1992. (Nursalim et al., 2023) Metode klasifikasi Support Vector Machine (SVM) merupakan algoritma yang memungkinkan pemisahan dua kelompok data dari dua kelas yang berbeda dengan optimal, dengan tujuan memaksimalkan jarak antara dua kelas melalui sebuah *hyperplane*. (Palisoa et al., 2023) metode merupakan metode yang sangat efektif untuk klasifikasi, menunjukkan hasil yang baik ketika digunakan dalam beragam domain. (Suryati & Aldino, 2023)

HASIL DAN PEMBAHASAN

Data Awal

Dalam tahap ini, digunakan data dari 2.000 komentar yang berasal dari berbagai channel YouTube yang membahas pengungsi Rohingya, seperti tvOneNews, KOMPASTV, Liputan6, METRO TV, dan CNN Indonesia. Proses pengumpulan data dilakukan melalui teknik web

crawling menggunakan bahasa Python yang tersedia di Google Colab. Data awal dapat dilihat pada sampel data dalam Tabel 1.

Tabel 1. Data awal

Author	Published_at	Update_at	Like_count	Text
@mamihooh560	2024-05-17T14:53:25Z	2024-05-17T14:53:25Z	0	Kenapa hrs di usir...dia kan ga ngerepotin wrga ...yng kasih mkn pemerintah biar z yng pnting aturan2 pemerintah tegas...
@Sakirsakir-zc5pm	2024-05-14T12:02:54Z	2024-05-14T12:02:54Z	0	Pengungsi aman, lebih lebih penduduk setempat harus aman.
@UripTarno	2024-05-13T16:02:25Z	2024-05-13T16:02:25Z	0	Tolak berbahaya takut bawa penyakit moral.susah ga ada obatnya.
@UripTarno	2024-05-13T15:58:42Z	2024-05-13T15:58:42Z	0	Saya rakyat indonesia menolak pengungsi rohingya takut rusak moral bangsa indonesia .sangat berbahaya buat bangsa indonesia.
@user-ky8ju9mg7w	2024-05-13T13:32:19Z	2024-05-13T13:32:19Z	0	Walau demi kemanusiaan klu bikin kusut buat nkri mending diusir 🤔🤔🤔🤔
@ashar_aryadi	2024-04-25T03:37:12Z	2024-04-25T03:37:12Z	0	Udah pulangkan aja orang rohingia itu itu bukan tanggung jawab orang Indonesia

Pelabelan

Setelah melakukan pengumpulan *dataset* melalui *crawling* dan pembersihan data, termasuk eliminasi komentar yang berulang dan tidak relevan dengan berita tentang pengungsi Rohingya, dilakukan tahap pelabelan untuk menentukan sentimen positif atau negatif dari setiap komentar. Contoh hasil pelabelan dapat dilihat pada Tabel 2. Proses ini penting untuk menganalisis respons dan pendapat publik terhadap isu tersebut secara lebih mendalam. Dengan demikian, kami dapat memahami lebih baik bagaimana persepsi masyarakat terhadap masalah pengungsi Rohingya. Langkah selanjutnya adalah menganalisis data yang telah dilabeli ini untuk mengeksplorasi tren sentimen dan pola opini yang muncul dari komunitas *online*.

Tabel 2. Case Folding

Text	Sentimen
Kenapa hrs di usir...dia kan ga ngerepotin wrga ...yng kasih mkn pemerintah biar z yng pnting aturan2 pemerintah tegas...	Positif
Pengungsi aman, lebih lebih penduduk setempat harus aman.	Positif
Tolak berbahaya takut bawa penyakit moral.susah ga ada obatnya.	Negative
Saya rakyat indonesia menolak pengungsi rohingya takut rusak moral bangsa indonesia .sangat berbahaya buat bangsa indonesia.	Negative
Walau demi kemanusiaan klu bikin kusut buat nkri mending diusir 🤔🤔🤔🤔	Negatif
Udah pulangkan aja orang rohingia itu itu bukan tanggung jawab orang Indonesia	Negatif

Pre-processing

Sebelum memasuki proses klasifikasi menggunakan algoritma data mentah yang diperoleh harus melalui proses *pre-processing*. Data mentah yang diperoleh melalui tahapan *crawling* masih bersifat tidak terstruktur, sehingga proses *pre-processing* diperlukan agar data mentah tersebut dapat diproses. Proses *pre-processing* merupakan proses yang bertujuan untuk menjadikan data mentah kedalam format yang lebih mudah dimengerti. Dalam proses ini terdapat beberapa tahapan yakni:

a) *Case Folding*

Penggunaan *Case Folding* adalah proses mengubah semua karakter menjadi huruf kecil. Contoh hasil dari *Case Folding* dapat dilihat dalam Tabel 3. Metode ini berguna untuk mengstandarisasi teks dan memudahkan perbandingan. Dengan *Case Folding*, perbedaan huruf besar dan kecil diabaikan, sehingga memungkinkan konsistensi dalam analisis teks. Hasil dari proses ini adalah teks yang lebih seragam dan mudah diolah secara komputasional. *Case Folding* mempermudah proses pencarian dan manipulasi teks dalam berbagai aplikasi, mulai dari pemrosesan bahasa alami hingga pemrograman.

Tabel 3. *Case Folding*

<i>Text</i>	<i>Case Folding</i>
Kenapa hrs di usir...dia kan ga ngerepotin wrga ...yng kasih mkn pemerintah biarin z yng pnting aturan2 pemerintah tegas...	kenapa hrs di usir...dia kan ga ngerepotin wrga ...yng kasih mkn pemerintah biarin z yng pnting aturan2 pemerintah tegas...
Pengungsi aman, lebih lebih penduduk setempat harus aman.	pengungsi aman, lebih lebih penduduk setempat harus aman.
Tolak berbahaya takut bawa penyakit moral.susah ga ada obatnya.	tolak berbahaya takut bawa penyakit moral.susah ga ada obatnya
Saya rakyat indonesia menolak pengungsi rohingya takut rusak moral bangsa indonesia .sangat berbahaya buat bangsa indonesia.	saya rakyat indonesia menolak pengungsi rohingya takut rusak moral bangsa indonesia .sangat berbahaya buat bangsa indonesia.
Walau demi kemanusiaan klu bikin kusut buat nkri mending diusir 😊😊😊😊	walau demi kemanusiaan klu bikin kusut buat nkri mending diusir 😊😊😊😊
Udah pulangkan aja orang rohingia itu itu bukan tanggung jawab orang Indonesia	udah pulangkan aja orang rohingia itu itu bukan tanggung jawab orang indonesia

b) *Cleaning*

Pada tahap *cleaning* atau pembersihan, teks dibersihkan dari kata-kata yang tidak relevan, tanda baca seperti *delimiter*, dan simbol-simbol lain yang tidak diperlukan. Contoh hasil dari proses *cleaning* dapat ditemukan dalam Tabel 4. Langkah ini penting dalam pra-pemrosesan teks karena membantu menyederhanakan konten dan meningkatkan kualitas analisis. *Cleaning* juga membantu mengurangi noise atau gangguan dalam data teks sehingga memperoleh hasil yang lebih akurat dalam pemrosesan berikutnya. Dengan membersihkan teks, informasi yang tersimpan menjadi lebih terfokus dan mudah dipahami oleh algoritma analisis teks.

Tabel 4. *Cleaning*

<i>Text</i>	<i>Cleaning</i>
kenapa hrs di usir...dia kan ga ngerepotin wrga ...yng kasih mkn pemerintah biarin z yng pnting aturan2 pemerintah tegas...	kenapa hrs di usir dia kan ga ngerepotin wrga yng kasih mkn pemerintah biarin z yng pnting aturan pemerintah tegas
pengungsi aman, lebih lebih penduduk setempat harus aman.	pengungsi aman lebih lebih penduduk setempat harus aman
tolak berbahaya takut bawa penyakit moral.susah ga ada obatnya	tolak berbahaya takut bawa penyakit moral susah ga ada obatnya
saya rakyat indonesia menolak pengungsi rohingya takut rusak moral bangsa indonesia .sangat berbahaya buat bangsa indonesia.	saya rakyat indonesia menolak pengungsi rohingya takut rusak moral bangsa indonesia sangat berbahaya buat bangsa indonesia
walau demi kemanusiaan klu bikin kusut buat nkri mending diusir 😊😊😊😊	walau demi kemanusiaan klu bikin kusut buat nkri mending diusir
udah pulangkan aja orang rohingia itu itu bukan tanggung jawab orang indonesia	udah pulangkan aja orang rohingia itu itu bukan tanggung jawab orang indonesia

c) *Tokenizing*

Selanjutnya, tahapan *tokenizing* dilakukan untuk memecah kalimat menjadi beberapa kata yang terpisah. Contoh hasil *tokenizing* dapat ditemukan dalam Tabel 5. Proses ini penting karena memungkinkan analisis lebih lanjut pada tingkat kata-kata, memungkinkan pemahaman yang lebih mendalam terhadap teks yang sedang diproses. Dengan memisahkan teks menjadi unit-unit kata, analisis lebih lanjut seperti klasifikasi atau analisis sentimen dapat dilakukan secara lebih efektif. Tahapan *tokenizing* membantu dalam mengurai teks secara terstruktur, memungkinkan pemrosesan lebih lanjut dengan algoritma pemrosesan bahasa alami atau analisis teks lainnya.

Tabel 5. *Tokenizing*

<i>Text</i>	<i>Cleaning</i>
kenapa hrs di usir dia kan ga ngerepotin wrga yng kasih mkn pemerintah biarin z yng pnting aturan pemerintah tegas	kenapa, hrs, di, usir, dia, kan, ga, ngerepotin, wrga, yng, kasih, mkn, pemerintah, biarin, z, yng, pnting, aturan, pemerintah, tegas
pengungsi aman lebih lebih penduduk setempat harus aman	pengungsi, aman, lebih, lebih, penduduk, setempat, harus, aman
tolak berbahaya takut bawa penyakit moral susah ga ada obatnya	tolak, berbahaya, takut, bawa, penyakit, moral, susah, ga, ada, obatnya
saya rakyat indonesia menolak pengungsi rohingya takut rusak moral bangsa indonesia sangat berbahaya buat bangsa indonesia	saya, rakyat, Indonesia, menolak, pengungsi rohingya, takut, rusak, moral, bangsa, indonesia, sangat, berbahaya, buat, bangsa, indonesia
walau demi kemanusiaan klu bikin kusut buat nkri mending diusir	walau, demi, kemanusiaan, klu, bikin, kusut, buat, nkri, mending, diusir
udah pulangkan aja orang rohingia itu itu bukan tanggung jawab orang indonesia	udah, pulangkan, aja, orang, rohingia, itu, itu, bukan, tanggung, jawab, orang, indonesia

d) *Stopword*

Tahap *stopword* berperan dalam menghilangkan kata-kata yang memiliki tingkat informasi yang rendah dari teks. Contoh hasil dari proses *stopword* dapat dilihat dalam Tabel 6. Langkah ini esensial dalam pra-pemrosesan teks karena membantu mengurangi noise dan meningkatkan relevansi informasi yang tersedia untuk analisis lebih lanjut. Dengan menghapus *stopword*, fokus analisis lebih tertuju pada kata-kata yang membawa makna penting dalam teks, sehingga menghasilkan hasil yang lebih relevan dan akurat dalam proses analisis teks.

Tabel 6. *Stopword*

<i>Text</i>	<i>Stopword</i>
kenapa, hrs, di, usir, dia, kan, ga, ngerepotin, wrga, yng, kasih, mkn, pemerintah, biarin, z, yng, pnting, aturan, pemerintah, tegas	kenapa di usir ga ngerepotin wrga yng kasih mkn pemerintah yng pnting aturan pemerintah tegas
pengungsi, aman, lebih, lebih, penduduk, setempat, harus, aman	pengungsi aman penduduk setempat harus aman
tolak, berbahaya, takut, bawa, penyakit, moral, susah, ga, ada, obatnya	tolak berbahaya bawa penyakit moral
saya, rakyat, Indonesia, menolak, pengungsi rohingya, takut, rusak, moral, bangsa, indonesia, sangat, berbahaya, buat, bangsa, indonesia	rakyat Indonesia menolak pengungsi rohingya rusak bangsa indonesia berbahaya bangsa indonesia
walau, demi, kemanusiaan, klu, bikin, kusut, buat, nkri, mending, diusir	kemanusiaan bikin kusut nkri diusir
udah, pulangkan, aja, orang, rohingia, itu, itu, bukan, tanggung, jawab, orang, indonesia	pulangkan rohingia bukan tanggung jawab indonesia

e) Normalisasi

Normalisasi bertujuan untuk mengubah kalimat yang tidak standar menjadi kalimat yang standar. Contoh hasil dari proses normalisasi dapat ditemukan dalam Tabel 7. Langkah ini penting dalam menghasilkan konsistensi dalam teks, sehingga mempermudah analisis dan pemrosesan lebih lanjut. Dengan melakukan normalisasi, variasi dalam ejaan atau format teks dapat diseragamkan, memudahkan pemahaman dan pengolahan teks oleh algoritma komputer. Proses ini memungkinkan pemrosesan lebih lanjut dengan menggunakan aturan yang sudah ditentukan secara konsisten.

Tabel 7. Normalisasi

<i>Text</i>	<i>Normalisasi</i>
kenapa di usir ga ngerepotin wrga yng kasih mkn pemerintah yng pnting aturan pemerintah tegas	kenapa di usir tidak ngerepotin warga yang memberi makan pemerintah yang penting aturan pemerintah tegas
pengungsi aman penduduk setempat harus aman	pengungsi aman penduduk setempat harus aman
tolak berbahaya bawa penyakit moral rakyat Indonesia menolak pengungsi rohingya rusak bangsa indonesia berbahaya bangsa indonesia	tolak berbahaya bawa penyakit moral rakyat Indonesia menolak pengungsi rohingya rusak bangsa indonesia berbahaya bangsa indonesia
kemanusiaan bikin kusut nkri diusir pulangkan rohingya bukan tanggung jawab indonesia	kemanusiaan bikin kusut nkri diusir pulangkan rohingya bukan tanggung jawab indonesia

f) Stemming

Tahap *stemming* memiliki tujuan untuk menghapus imbuhan pada awal atau akhir kata sehingga kata tersebut menjadi kata dasar. Contoh hasil stemming dapat dilihat dalam Tabel 8. Proses ini memperoleh kata dasar dari variasi kata yang ada, membantu dalam mengurangi kompleksitas dan memperbaiki konsistensi dalam analisis teks. Dengan mengaplikasikan stemming, kemungkinan kemunculan kata yang serupa namun berbeda bentuknya dapat diperkecil, sehingga meningkatkan akurasi dan efisiensi dalam pemrosesan teks.

Tabel 8. Stemming

<i>Text</i>	<i>Stemming</i>
kenapa di usir ga ngerepotin wrga yng kasih mkn pemerintah yng pnting aturan pemerintah tegas	kenapa usir tidak repot warga memberi makan pemerintah penting atur pemerintah tegas
pengungsi aman penduduk setempat harus aman	pengungsi aman penduduk setempat harus aman
tolak berbahaya bawa penyakit moral rakyat Indonesia menolak pengungsi rohingya rusak bangsa indonesia berbahaya bangsa indonesia	tolak bahaya bawa penyakit moral rakyat Indonesia tolak pengungsi rohingya rusak bangsa indonesia bahaya bangsa indonesia
kemanusiaan bikin kusut nkri diusir pulangkan rohingya bukan tanggung jawab indonesia	kemanusiaan bikin kusut nkri usir pulang rohingya bukan tanggung jawab indonesia

TF-IDF

Langkah selanjutnya adalah pembobotan TF-IDF yang dilakukan menggunakan *dataset* yang telah melalui proses *pre-processing* dan pelabelan pada tahap sebelumnya. Tujuan dari pembobotan ini adalah menentukan nilai terhadap kata dalam suatu dokumen dan mengidentifikasi kata-kata kunci yang mempengaruhi dokumen tersebut. Hasil sampel dalam pembobotan TF-IDF bisa dilihat pada Tabel 9.

Tabel 9. TF-IDF

Text	TF-IDF
kenapa usir tidak repot warga memberi makan pemerintah penting atur pemerintah tegas	"kenapa": 1.792, "usir": 1.099, "tidak": 1.792, "repot": 1.792, "warga": 1.792, "memberi": 1.792, "makan": 1.792, "pemerintah": 2.198, "penting": 1.792, "atur": 1.792, "tegas": 1.792
pengungsi aman penduduk setempat harus aman	"pengungsi": 1.099, "aman": 2.198, "penduduk": 1.792, "setempat": 1.792, "harus": 1.792
tolak bahaya bawa penyakit moral	"tolak": 1.099, "bahaya": 1.099, "bawa": 1.792, "penyakit": 1.792, "moral": 1.792
rakyat Indonesia tolak pengungsi rohingya rusak bangsa indonesia bahaya bangsa indonesia	"rakyat": 1.792, "Indonesia": 1.099, "tolak": 1.099, "pengungsi": 1.099, "rohingya": 1.099, "rusak": 1.792, "bangsa": 2.198, "indonesia": 1.398, "bahaya": 1.099
kemanusiaan bikin kusut nkri usir	"kemanusiaan": 1.792, "bikin": 1.792, "kusut": 1.792, "nkri": 1.792, "usir": 1.099
pulang rohingya bukan tanggung jawab indonesia	"pulang": 1.792, "rohingya": 1.099, "bukan": 1.792, "tanggung": 1.792, "jawab": 1.792, "indonesia": 0.699

Klasifikasi SVM

Setelah *dataset* melalui tahapan *pre-processing*, pelabelan, dan penerapan TF-IDF, langkah berikutnya adalah menggunakan algoritma Support Vector Machine (SVM) untuk klasifikasi. SVM dipilih untuk menganalisis sentimen, memisahkan dokumen teks ke dalam kelas positif atau negatif berdasarkan pembobotan fitur. Data akan dibagi menjadi tiga skenario dengan rasio training dan testing yang berbeda: 70:30, 80:20, dan 90:10. Dengan variasi rasio ini, kinerja model SVM diuji dalam memprediksi sentimen dengan akurasi yang berbeda, memungkinkan pemilihan rasio optimal untuk aplikasi nyata. Hasil *training* dari ketiga rasio tersedia dalam Tabel 11, 12 dan 13 memberikan wawasan tentang performa model dalam konteks pembagian data yang telah ditentukan.

Tabel 10. Skenario data training dan data testing

Skenario	Rasio	Data Training	Data Testing
1	70:30	1400	600
2	80:20	1600	400
3	90:10	1800	200

Tabel 11. Training Rasio 70:30

Akurasi	Precision	Recall	F1-Score	Support
Positive	0.00	0.00	0.00	73
Negative	0.74	1.00	0.85	207
Accuracy			0.74	280
Macro avg	0.37	0.50	0.43	280
Weighted avg	0.55	0.74	0.63	280

Tabel 12. Training Rasio 80:20

Akurasi	Precision	Recall	F1-Score	Support
Positive	0.00	0.00	0.00	79
Negative	0.75	1.00	0.86	241
Accuracy			0.75	320
Macro avg	0.38	0.50	0.43	320
Weighted avg	0.57	0.74	0.65	320

Tabel 13. Training Rasio 90:10

Akurasi	<i>Precision</i>	<i>Recall</i>	F1-Score	Support
Positive	1.00	0.01	0.02	84
Negative	0.77	1.00	0.87	276
Accuracy			0.77	360
Macro avg	0.88	0.51	0.45	360
Weighted avg	0.82	0.77	0.67	360

Evaluasi

Berdasarkan hasil training dengan rasio data 70:30, 80:20, dan 90:10, terlihat bahwa model memiliki performa yang bervariasi tergantung pada rasio data yang digunakan. Pada rasio 70:30, model menunjukkan akurasi sebesar 74% dengan *precision* dan *recall* untuk kelas positif sangat rendah (0.00) dan sangat tinggi untuk kelas negatif, menunjukkan model hampir tidak mampu mengidentifikasi kelas positif dengan benar. Pada rasio 80:20, akurasi sedikit meningkat menjadi 75%, namun *precision* dan *recall* untuk kelas positif tetap rendah, menandakan masalah ketidakseimbangan kelas masih ada.

Pada rasio 90:10, akurasi keseluruhan meningkat menjadi 77%, tetapi meskipun *precision* untuk kelas positif meningkat menjadi 1.00, *recall* sangat rendah (0.01), menunjukkan model masih kesulitan mengidentifikasi kelas positif dengan benar. Secara keseluruhan, model lebih baik dalam mengklasifikasikan kelas negatif dibandingkan kelas positif. Ketidakseimbangan kelas dalam *dataset* tampaknya menjadi penyebab utama rendahnya performa pada kelas positif. Oleh karena itu, diperlukan metode penanganan ketidakseimbangan data seperti oversampling kelas minoritas, undersampling kelas mayoritas, atau penggunaan algoritma yang lebih robust terhadap ketidakseimbangan kelas untuk meningkatkan performa model.

Kesimpulan

Hasil klasifikasi menggunakan algoritma SVM dengan rasio data 70:30, 80:20, dan 90:10 menunjukkan model lebih baik dalam mengklasifikasikan kelas negatif dibandingkan kelas positif. Meskipun akurasi meningkat dengan rasio data yang lebih tinggi, model masih kesulitan mengidentifikasi kelas positif dengan benar karena ketidakseimbangan data. Untuk meningkatkan performa, diperlukan metode penanganan ketidakseimbangan data.

Saran

Untuk meningkatkan performa model, disarankan untuk menangani ketidakseimbangan data dengan metode seperti oversampling kelas minoritas, undersampling kelas mayoritas, atau menggunakan algoritma yang lebih robust seperti SMOTE. Evaluasi model menggunakan metrik seperti AUC atau *Precision-Recall* Curve juga dapat memberikan gambaran yang lebih baik tentang performa terhadap kelas minoritas.

Ucapan Terimakasih

Saya ingin mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam proses akademik dan penelitian ini. Terima kasih atas waktu, tenaga, dan pemikiran yang telah Anda curahkan untuk membantu serta memberikan bimbingan yang sangat berharga bagi kami.

REFERENSI

- Andzar, A. M. (2023). *Penerapan Text mining Dalam Klasifikasi Judul Skripsi Yang Diusulkan Mahasiswa Menggunakan Metode Naïve Bayes*. 6, 184–196.
- Di, A., Menggunakan, P. A., & Naïve, M. (2024). *G-Tech : Jurnal Teknologi Terapan*. 8(2), 847–856.
- Izzah Nur Masyithoh, & Ivo Novitaningtyas. (2021). Pengaruh Digital Marketing Terhadap Minat Beli Konsumen Pada Marketplace Tokopedia. *Jurnal Manajemen & Bisnis Kreatif*, 7(1). <https://doi.org/10.36805/manajemen.v7i1.1951>

- Kamilla, A. C., Priyani, N., Priskila, R., & Pranatawijaya, V. H. (2024). *Analisis Sentimen Film Agak Laen Dengan Kecerdasan Buatan : Text Mining Metode Naïve Bayes Classifier*. 8(3), 2923–2928.
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). *Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine*. 2(11), 4704–4713.
- Maulana, B. A., & Fahmi, M. J. (2024). *Sentiment Analysis of Pluang Applications With Naive Bayes and Support Vector Machine (SVM) Algorithm Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)*. 4(April), 375–384.
- Muis, A., & Muhammad, F. (n.d.). *Pelatihan Text mining Menggunakan Bahasa Pemrograman Python*. 36–46.
- Nursalim, A., Novita, R., Systems, I., & Program, S. (2023). *Sentiment Analysis Of Comments On Google Play Store , Twitter And Youtube To The Mypertamina Application With Support Vector Analisis Sentimen Komentar Pada Google Play Store , Twitter Dan Youtube Terhadap Aplikasi Mypertamina Dengan Support*. 4(6), 1305–1312.
- Palisoa, N. F., Sinay, L. J., Matdoan, M. Y., Yudistira, Y., & Bakarbesy, L. (2023). Penerapan Support Vector Machine (SVM) Untuk Klasifikasi Kabupaten Tertinggal Di Provinsi Maluku. *PARAMETER: Jurnal Matematika, Statistika Dan Terapannya*, 2(02), 79–86. <https://doi.org/10.30598/parameterv2i02pp79-86>
- Saputra, Y. A. (2024). *Sentimen Mahasiswa Terhadap Materi Pembelajaran Komputer : Studi Kasus Dan Analisis*. 1(3), 1–16.
- Sentimen, A., Ulasan, P., Maxim, A., Play, G., Dengan, S., Ramadhan, R., Sains, F., Teknologi, D. A. N., Islam, U., Sultan, N., & Kasim, S. (2023). *Tugas akhir*.
- Suryati, E., & Aldino, A. A. (2023). *Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)*. 4(1), 96–106.
- Sutresno, S. A. (2023). *Analisis Sentimen Masyarakat Indonesia Terhadap Dampak Penurunan Global Sebagai Akibat Resesi di Twitter*. 4(4), 1959–1966. <https://doi.org/10.47065/bits.v4i4.3149>
- Widianto, E. (2021). Pemanfaatan Media Pembelajaran Berbasis Teknologi Informasi. *Journal of Education and Teaching*, 2(2), 213. <https://doi.org/10.24014/jete.v2i2.11707>
- Zelina, N., & Afyati, A. (2024). *Analisis Sentimen Ulasan Pengguna Aplikasi M- Banking Menggunakan Algoritma Support Vector Machine dan Decision Tree*. 7(1), 31–37.