Terakreditasi Peringkat 4
Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi, No. 225/E/KPT/2022
Masa berlaku dari Volume 8 Nomor 1 Tahun 2022 sampai dengan Volume 12 Nomor 2 Tahun 2026

Published online at: **https://ejournal.raharja.ac.id/**

JOURNAL CERITA:
Creative Education of Research in Information Technology and Artificial informatics

Vol. 12 No. 2 (2026) 187-198 e-ISSN: 2655 - 2574

Analisis Performa InSet Lexicon Pada Pelabelan Sentimen Pengungsi Rohingya Menggunakan SVM

Rafi Aditya Mahendra[1], Yisti Vita Via[*2], Chrystia Aji Putra[3]
[1,2,3]Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia
Email: rafipersonalporto@gmail.com[1]; **yistivia.if@upnjatim.ac.id[2]**; ajiputra@upnjatim.ac.id[3]

Mahendra, R. A., Via, Y. V., & Putra, C. A. (2026). Analisis Performa InSet Lexicon Pada Pelabelan Sentimen Pengungsi Rohingya Menggunakan SVM. *Journal Cerita: Creative Education of Research in Information Technology and Artificial Informatics*, 12(2), 187-198

DOI: https://doi.org/10.33050/cerita.v12i2.3706

ABSTRAK

Analisis sentimen adalah studi yang mengevaluasi opini, sentimen, penilaian, sikap, dan emosi seseorang terhadap entitas tertentu seperti masalah atau peristiwa. Salah satu pendekatan umum adalah metode lexicon-based, di mana teks dianalisis menggunakan kamus sentimen seperti InSet Lexicon yang mengelompokkan kata ke dalam kategori positif atau negatif. Dataset yang digunakan dalam penelitian ini adalah tweet dari aplikasi X (Twitter) mengenai pengungsi Rohingya di Indonesia. Klasifikasi data dilakukan dengan algoritma Support Vector Machine (SVM). Hasil dari pelabelan menggunakan InSet Lexicon ditemukan 886 sentimen positif, 2906 sentimen negatif, dan 353 sentimen netral dengan rata-rata nilai akurasi yang dihasilkan dari seluruh pengujian mencapai 84.9%.

Kata kunci: Analisis Sentimen, Support Vector Machine, InSet Lexicon

ABSTRACT

Sentiment analysis is a study that evaluates a person's opinions, sentiments, judgments, attitudes, and emotions towards certain entities such as issues or events. One common approach is the lexicon-based method, where text is analyzed using a sentiment dictionary such as InSet Lexicon that groups words into positive or negative categories. The dataset used in this research is tweets from application X (Twitter) regarding Rohingya refugees in Indonesia. Data classification is done with the Support Vector Machine (SVM) algorithm. The results of labeling using InSet Lexicon found 886 positive sentiments, 2906 negative sentiments, and 353 neutral sentiments with an average accuracy value generated from all tests reaching 84.9%.

Keywords: *Sentiment Analysis, Support Vector Machine, InSet Lexicon*

Received: 20-02-2026 | Accepted: 20-06-2026 | Published Online: 08-08-2026

I. PENDAHULUAN

Kisah pengungsi Rohingnya yang mencari simpati dan bantuan ke berbagai negara di Asia Tenggara menjadi isu krisis kemanusiaan yang menarik perhatian dunia internasional pada tahun 2023. Isu tersebut juga menarik perhatian masyarakat Indonesia sejak kedatangan 2000 pengungsi Rohingnya di Aceh. Penolakan keras dilakukan oleh masyarakat Aceh yang sebelumnya mengalami kejadian tidak mengenakan karena menerima para pengungsi Rohingnya. Para pengungsi Rohingnya yang diterima di Aceh justru melarikan diri dari tempat pengungsian, tidak menghormati norma agama dan norma sosial masyarakat setempat, serta melakukan tindakan kriminal yang mengganggu keamanan dan kondusivitas masyarakat Aceh (R. Baidi, 2024).

Selain pernyataan penolakan dari masyarakat Aceh terhadap pengungsi Rohingnya, masyarakat Indonesia dari wilayah lain pun turut memberikan pendapat mereka dalam media sosial. Media sosial merupakan media yang dapat digunakan pengguna dalam rangka mempresentasikan dirinya, berinteraksi, bekerja sama, berbagai informasi, maupun berinteraksi dengan pengunjung lainnya (Krisdiyanto, 2021). Melalui media sosial, masyarakat mampu mengekspresikan pendapatnya terkait suatu hal dan diskusi secara publik, termasuk terkait orang etnis Rohingnya yang mencari simpati di Indonesia. X atau lebih dikenal dahulu dengan sebutan Twitter merupakan satu dari *platform* media sosial yang memiliki pengaruh besar di Indonesia. Hal ini terbukti dari peringkatnya yang berada pada urutan ke-5 dengan 30,2% pengguna berdasarkan survei Katadata Insight Center (KIC) pada tahun 2023 (Muhammad, 2023). Melalui X, pengguna dapat mengunggah konten sesuai keinginan, seperti opini maupun *emoticon* yang dapat menjadi data untuk menganalisis suatu *trend* atau topik tertentu (Musfiroh et al., 2021). Satu dari topik yang ramai diperbincangankan di X adalah pengungsi Rohingnya di Indonesia tahun 2023. Banyak pengguna X yang membagikan opini mereka melalui *tweet*, baik opini positif maupun negatif. Penelitian ini bertujuan melakukan analisis sentimen dalam *tweet* yang membahas pengungsi Rohingnya di Indonesia pada tahun 2023.

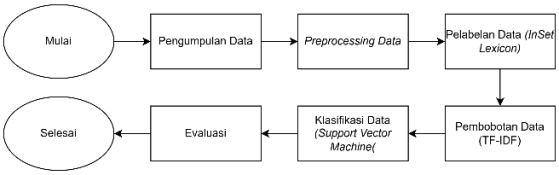
Analisis sentimen merupakan sebuah studi yang menganalisis pendapat, sentimen, penilaian, sikap, evaluasi, dan emosi seseorang terhadap entitas seperti sebuah produk, layanan, organisasi, individual, masalah, topik atau peristiwa, dan semacamnya (Yusuf et al., 2023). Analisis sentimen dapat dilakukan dengan pendekatan *supervised learning* dan pendekatan berbasis *lexicon*, serta dapat menggunakan berbagai algoritma seperti *Naïve Bayes*, *Maxium Entropy*, dan *Support Vector Machne*. Adapun yang akan digunakan dalam penelitian ini adalah pendekatan dengan metode *lexicon-based* dengan algoritma *Support Vector Machine* (SVM). Metode *lexicon-based* akan menganalisis teks berdasarkan kamus sentimen yang mengelompokkan kata-kata ke dalam kategori sentimen positif atau negatif. *Lexicon -based* menggunakan kamus sebagai dasar bahasa atau leksikal (Mulyadi & Lestari, 2022). Dalam kamus *lexicon*, kata-kata didentifikasi atau dibandingkan berdasarkan polaritasnya (Kharde & Sonawane, 2016). Adapun kamus sentimen yang digunakan adalah InSet (Indonesia *Sentiment*) Lexicon. Kamus tersebut berisi daftar kata yang mengandung 3609 kata yang dianggap positif dan 6609 kata yang dianggap negatif dengan bobot nilai berkisar -5 hingga +5 (Prasetya et al., 2021). Sedangkan algoritma Support Vector Machine (SVM) adalah model *Supervised Learning* yang digunakan untuk analisis data dengan tujuan melakukan analisis regresi dan klasifikasi (Styawati et al., 2021). *Support Vector Machine* (SVM) memiliki tujuan utama untuk menemukan *hyperplane* atau batas keputusan yang memisahkan jarak antar kelas terbaik (Styawati et al., 2021). Proses pencari letak *hyperplane* terbaik dilakukan pada proses pelatihan model dan hasil terbaik akan digunakan untuk pengujian model (Wulandari, 2020).

Penelitian ini akan mengambil data berupa sentimen dari *tweet* pengguna media sosial X yang membahas pengungsi Rohingnya di Indonesia pada tahun 2023. Dengan proses analisisnya mengelompokan data menjadi dua kategori sentimen, yaitu positif dan negatif. Menggunakan InSet Lexicon untuk proses pelabelan data serta algoritma *Support Vector Machine* (SVM) untuk proses klasifikasi data.

II. METODE PENELITIAN

Dalam melakukan penelitian, peneliti melalui beberapa tahapan, seperti pengumpulan data, *preprocessing data*, pelabelan data, pembobotan data, klasifikasi data, serta evaluasi

hasil. Berikut ini adalah diagram alur penelitian secara umum yang dilakukan oleh peneliti.



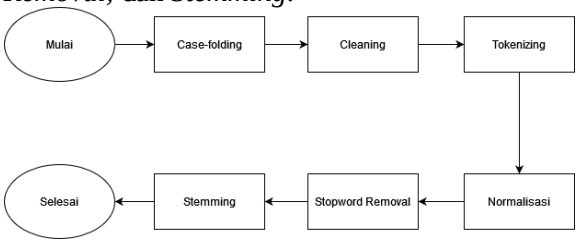
Gambar 1. Alur Penelitian
Sumber: diolah dari data primer

A. Pengumpulan Data

Pengumpulan atau akuisisi atau crawling data dilakukan dengan mengambil data berupa tweet pada aplikasi X dengan library "tweet-harvest". Tweet yang diambil adalah tweet yang membahas mengenai pengungsi Rohingnya di Indonesia pada tahun 2023. Hasil pengumpulan data akan berisikan informasi kapan tweet dibuat, siapa pembuatnya, isi tweet atau konten, bahasa yang digunakan, serta link url dari tweet yang diunggah.

B. Preprocessing Data

Setelah dilakukan pengumpulan data maka data akan diproses untuk dibersihkan agar formatnya sesuai dengan kebutuhan sistem. Tahap yang akan dilalui meliputi *Case-Folding, Cleaning, Tokenizing,* Normalisasi, *Stopword Removal*, dan *Stemming*.



Gambar 2. Alur Preprocessing Data
Sumber: diolah dari data primer

1. *Case-Folding*
   Pada tahap ini, semua huruf dalam teks pada *tweet* akan diubah menjadi huruf kecil demi menjaga konsistensi dan konsolidasi teks.
2. *Cleaning*
   Tahap ini dilakukan dengan tujuan untuk menghilangkan karakter yang tidak relevan atau tidak diinginkan dalam teks. Adapun karakter yang dimaksud adalah karakter khusus, tanda baca, dan angka. Jadi, hanya informasi yang relevan saja yang akan diproses untuk analisis sentimen.
3. *Tokenizing*
   Pada tahap ini, akan dilakukan pemecahan teks menjadi unit-unit yang lebih kecil atau "token". Teks yang tadinya berupa kalimat akan dipecah menjadi kata-kata individual atau frasa.
4. Normalisasi
   Selanjutnya, akan dilakukan pengubahan kata-kata tidak baku menjadi kata baku yang mencakup mengganti sinonim atau variasi kata dengan bentuk dasar untuk memudah pemahan dan analisis.
5. *Stopword Removal*
   Pada tahap ini, kata-kata umum yang tidak memiliki nilai arti yang tinggi dalam sentimen akan dihapuskan dari dataset teks yang akan dianalisis, agar sistem fokus pada kata kunci yang lebih relevan.
6. *Stemming*
   Langkah terakhir dalam *data preprocessing* akan dilakukan penghilangan imbuhan dari kata-kata yang ada pada dataset agar menjadi kata dasar. Hal ini dilakukan untuk mengurangi dimensi teks dan meningkatkan konsistensi untuk memperkuat analisis sentimen.

C. Pelabelan Data

Setelah proses *preprocessing data* selesai dilakukan maka selanjutnya akan dilakukan pelabelan data menggunakan *Inset Lexicon*. Perhitungan nilai sentimen dilakukan dengan menghitung bobot negatif dan positif pada setiap kata, kemudian total perhitungan bobot kalimat akan menghasilkan nilai untuk menentukan kalimat tersebut masuk ke dalam kategori sentimen positif atau negatif. Adapun kententuan dalam penentuan jenis sentimen adalah sebagai berikut.

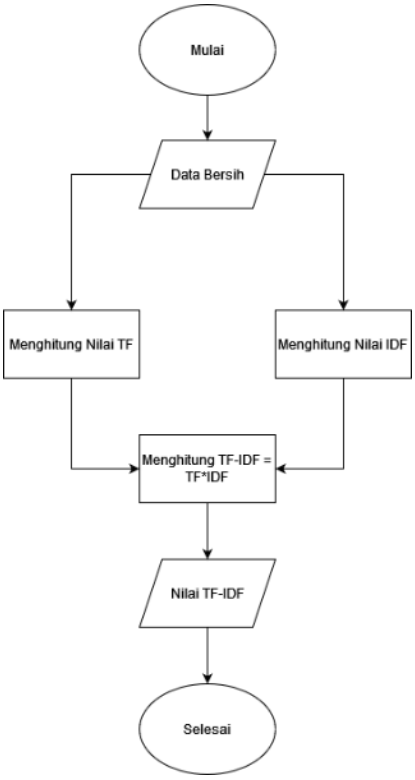
If sentiment score > 0	*then* Sentimen Positif
If sentiment score < 0	*then* Sentimen Negatif
If sentiment score $= 0$	*then* Sentimen Netral

Gambar 3. Ketentuan Pelabelan Data
Sumber: (Musfiroh et al., 2021)

D. Pembobotan Data

Tahap pembobotan data menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), di mana akan dilakukan penghitungan bobot kata pada setiap dokumen dalam kumpulan data. Langkah pertama adalah menghitung frekuensi kemunculan kata dalam dokumen yang disebut dengan *Term Frequency (TF)*. Langkah kedua menyeimbangkan bobot kata yang muncul di berbagai dokumen dengan membaginya terhadap nilai *Inverse Document Frequency* (IDF) yang mengukur seberapa jarang

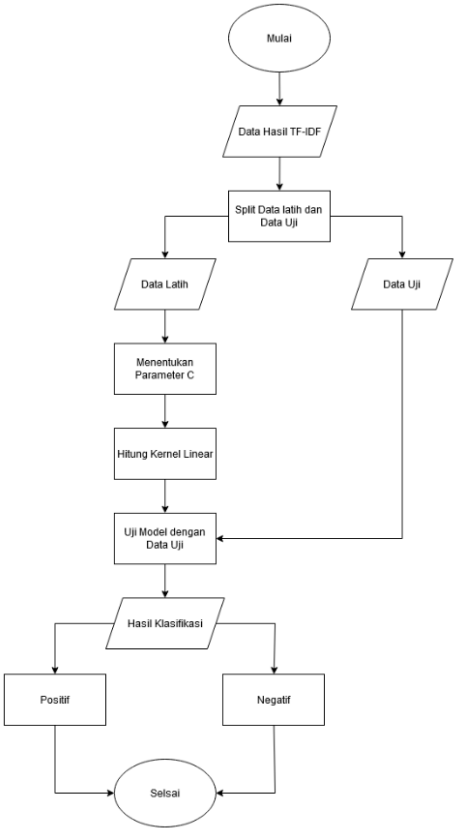
sebuah kata muncul pada sebuah dokumen dalam dataset. Hasil dari perhitungan ini akan menjadi bobot kata yang digunakan untuk analisis berikutnya. Melalui tahap ini, kata-kata yang sering muncul akan memiliki bobot yang lebih rendah dibandingkan kata-kata yang jarang muncul, namun khusus muncul di beberapa dokumen.



Gambar 4. Alur Pembobotan Data (TF-IDF)
Sumber: diolah dari data primer

E. Klasifikasi SVM

Setelah dilakukan pembobotan kata maka akan dilakukan klasifikasi data dengan algoritma *Support Vector Machine* (SVM) menggunakan kernel linear. Pertama, dilakukan pembagian data hasil dari perhitungan TF-IDF menjadi data latih dan data uji. Kedua, dilakukan inisialisasi *hyperparameter* C. Ketiga, dilakukan penghitungan mtariks kernel. Keempat, dilakukan pelatihan model untuk mencari model yang paling optimal. Tahap terakhir adalah pengujian yang dilakukan dengan data uji berdasarkan data yang dilatih sebelumnya. Berikut ini adalah detail alur dalam melakukan klasifikasi dengan *Support Vector Machine* (SVM).



Gambar 5. Alur Klasifikasi (SVM)
Sumber: diolah dari data primer

F. Evaluasi

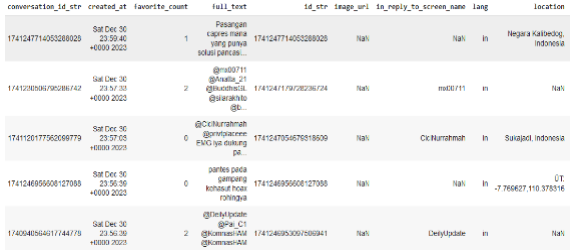
Setelah berhasil melakukan klasifikasi sentimen positif dan negatif, model akan diuji untuk melihat seberapa baik model bekerja dengan kondisi yang berbeda-beda. Adapun skenario pengujian yang akan dilakukan adalah membagi data menjadi 90% data latih dan 10% data uji, 80% data latih dan 20% data uji, dan 70% data latih dan 30% data uji. Kemudian, akan dievaluasi dengan *confusion matrix* untuk menghitung nilai akurasi, *precision, recall,* dan *f1*-score.

III. HASIL DAN PEMBAHASAN

Berikut ini adalah hasil penelitian yang didapatkan pada setiap alur penelitian yang dijabarkan pada bab ke-3.

A. Pengumpulan Data

Data yang dikumpulkan berupa *tweets* dari X sejak 1 Juli 2023 – 1 Desember 2023, dengan kata kunci "rohingya lang:id" yang berarti *tweet* tentang rohingya menggunakan bahasa Indonesia. Jumlah data yang terkumpul adalah 4222 *tweets*.



conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang	location
1741247714053288028	Sat Dec 30 23:59:40 +0000 2023	1	Pasangan capres mana yang punya solusi pancasi...	1741247714053288028	NaN	NaN	in	Negara Kabbedog, Indonesia
1741230506796286742	Sat Dec 30 23:57:33 +0000 2023	2	@mx00711 @Anatta_21 @BuddhisGL @silarakhito @b...	1741247119728236724	NaN	mx00711	in	NaN
1741120177562099779	Sat Dec 30 23:57:05 +0000 2023	0	@CiciNurrahmah @privtplaceee EMG iya dukung pa...	1741247054578618609	NaN	CiciNurrahmah	in	Sukajadi, Indonesia
1741246956608127988	Sat Dec 30 23:56:56 +0000 2023	0	pantes pada gampang kehasut hoax rohingya	1741246956608127988	NaN	NaN	in	0.7: -7.769827,110.378316
1740940564917447479	Sat Dec 30 23:56:39 +0000 2023	2	@DeilyUpdate @Pai_C1 @KomnasHAM @KomnasHAM men...	1741246853097586941	NaN	DeilyUpdate	in	NaN

Gambar 6. Hasil Pengumpulan Data
Sumber: diolah dari data primer

Sebelum dilanjutkan ke tahap *preprocessing data*, dilakukan penghapusan semua kolom pada dataset dengan hanya menyisakan kolom "full_text". Hal ini dilakukan agar sistem berfokus pada pemrosesan teks yang berisi inti konten atau *tweet*.

	full_text
0	Pasangan capres mana yang punya solusi pancasi...
1	@mx00711 @Anatta_21 @BuddhisGL @silarakhito @b...
2	@CiciNurrahmah @privtplaceee EMG iya dukung pa...
3	pantes pada gampang kehasut hoax rohingya
4	@DeilyUpdate @Pai_C1 @KomnasHAM @KomnasHAM men...
...	...
4217	@tarocotton @leethanizer @tanyarlfes Iya terus...
4218	@is_HANAA_29 @margianta @Refugees @UNHCRIndo T...
4219	Biasanya orang begini tak banyak pengalaman hi...
4220	@putri20297298 @hoo_mma @WvOuthoorn @margianta...
4221	@soundefire @M29___ @taeko0kie_ @rgantas Lu be...

Gambar 7. Hasil Penghapusan Kolom
Sumber: diolah dari data primer

B. Preprocessing Data

Pada tahap ini akan ditampilkan proses yang dilalui dalam mengubah format data yang sudah dikumpulkan agar sesuai kebutuhan sistem.

1. Case-Folding

Teks dalam *tweet* yang berisikan huruf kapital dan huruf kecil seperti pada gambar 7 akan disamakan menggunakan huruf kecil.

	teks
0	pasangan capres mana yang punya solusi pancasi...
1	@mx00711 @anatta_21 @buddhisgl @silarakhito @b...
2	@cicinurrahmah @privtplaceee emg iya dukung pa...
3	pantes pada gampang kehasut hoax rohingya
4	@deilyupdate @pai_c1 @komnasham @komnasham men...
5	@pai_c1 bagus kalau perlu naikkan semua pajak ...
6	status muslim yang disandang oleh orang² rohin...
7	rohingya ni camna nak hantar balik tempat diorang
8	@aburokoq @tanyakanrl rohingya sama capres itu...
9	@rodrichen @berlianidris kasus rohingys terdam...

Gambar 8. Hasil Proses Case-Folding
Sumber: diolah dari data primer

2. Cleaning

Dataset *tweet* pada gambar 8 akan dibersihkan dari elemen yang tidak diperlukan.

	hasil_cleaning
0	pasangan capres mana yang punya solusi pancasi...
1	tergantung sikon juga krna semua dibiasakan be...
2	dukung palestin tapi ngusir rohingya dasar fas...
3	pantes pada gampang kehasut hoax rohingya
4	mending tampung rohingya ditempat kalian duitn...
5	bagus kalau perlu naikkan semua pajak agar pem...
6	status muslim yang disandang oleh orang² rohin...
7	rohingya camna hantar balik tempat diorang
8	rohingya sama capres pembahasan umum aneh orang
9	kasus rohingys terdampar aceh puluhan tahun ba...

Gambar 9. Hasil Proses Cleaning
Sumber: diolah dari data primer

3. Tokenizing

Dataset pada gambar sebelumnya yang berbentuk kalimat akan dipecah menjadi "token".

	tokenizing
0	['pasangan', 'capres', 'mana', 'yang', 'punya'...
1	['tergantung', 'sikon', 'juga', 'krna', 'semua...
2	['dukung', 'palestin', 'tapi', 'ngusir', 'rohi...
3	['pantes', 'pada', 'gampang', 'kehasut', 'hoax...
4	['mending', 'tampung', 'rohingya', 'ditempat',...
5	['bagus', 'kalau', 'perlu', 'naikkan', 'semua'...
6	['status', 'muslim', 'yang', 'disandang', 'ole...
7	['rohingya', 'camna', 'hantar', 'balik', 'temp...
8	['rohingya', 'sama', 'capres', 'pembahasan', '...
9	['kasus', 'rohingys', 'terdampar', 'aceh', 'pu...

Gambar 10. Hasil Proses Tokenizing
Sumber: diolah dari data primer

4. Normalisasi

Kata pada dataset yang tidak baku akan diubah menjadi kata baku dengan acuan kamus dalam "colloquial-indonesian-lexicon.csv".

	hasil_normalisasi
0	['pasangan', 'capres', 'mana', 'yang', 'punya'...
1	['tergantung', 'sikon', 'juga', 'karena', 'sem...
2	['dukung', 'palestin', 'tapi', 'ngusir', 'rohi...
3	['pantas', 'pada', 'gampang', 'kehasut', 'hoax...
4	['mending', 'tampung', 'rohingya', 'ditempat',...
5	['bagus', 'kalau', 'perlu', 'naikkan', 'semua'...
6	['status', 'muslim', 'yang', 'disandang', 'ole...
7	['rohingya', 'camna', 'hantar', 'balik', 'temp...
8	['rohingya', 'sama', 'capres', 'pembahasan', '...
9	['kasus', 'rohingys', 'terdampar', 'aceh', 'pu...

Gambar 11. Hasil Proses Normalisasi
Sumber: diolah dari data primer

5. Stopword Removal

Dataset yang kata-katanya sudah baku akan dilakukan penghapusan untuk kata-kata yang tidak memiliki informasi penting.

	stopword_removal
0	['pasangan','capres','','','solusi','pancas...
1	['tergantung','sikon','','','dibiasakan','b...
2	['dukung','palestin','','ngusir','rohingya','d...
3	['','','gampang','kehasut','hoax','rohingya']
4	['mending','tampung','rohingya','ditempat','',...
5	['bagus','','','naikkan','','pajak','','pemban...
6	['status','muslim','','disandang','','orang','...
7	['rohingya','camna','hantar','','','diorang']
8	['rohingya','','capres','pembahasan','','aneh'...
9	['','rohingys','terdampar','aceh','puluhan','...

Gambar 12. Hasil Proses Stopword Removal
Sumber: diolah dari data primer

6. Stemming

Pada tahap ini dilakukan penghilangan imbuhan dari kata-kata yang ada pada dataset agar menjadi kata dasar.

	hasil_stemming
0	pasang capres solusi pancasila ihwal rohingya
1	gantung kondisi biasa dampak positif lingkung ...
2	dukung palestina ngusir rohingya dasar fasis m...
3	gampang hasut hoax rohingya
4	mending tampung rohingya tempat duit pakai pak...
5	bagus naik pajak bangun jalan rohingya makan m...
6	status muslim sandang orang rohingya orang ind...
7	rohingya camna hantar orang
8	rohingya capres bahas aneh orang
9	rohingys dampar aceh puluh terima tolong urus ...

Gambar 13. Hasil Proses Stemming
Sumber: diolah dari data primer

Langkah terakhir sebelum pelabelan data akan dilakukan penghapusan kalimat duplikat yang ada berada pada dataset.

	hasil_stemming	is_duplicate
0	pasang capres solusi pancasila ihwal rohingya	False
1	gantung sikon biasa dampak positif lingkung be...	False
2	dukung palestina usir rohingya dasar fasis mun...	False
3	gampang hasut hoax rohingya	False
4	mending tampung rohingya tempat duit pakai pak...	False
...	...	...
4140	hubung demo pengungsi mahasiswa kemarin kisruh...	False
4141	republika presiden joko widodo jokowi perintah...	False
4142	orang alam hidup ajar negara banjir warga asin...	False
4143	susah banget malaysia persekusi rohingya simpe...	False
4144	bego coba bayang rumah orang asing usir pergi ...	False

Gambar 14. Hasil Penghapusan Duplicate Data
Sumber: diolah dari data primer

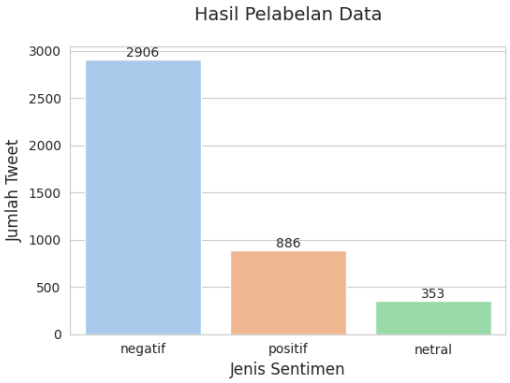
Total data yang berhasil didapatkan sebelumnya adalah 4222. Setelah melalui proses penghapusan data yang memiliki duplikat, jumlahnya mejadi 4145. Data tersebut yang akan diproses ke tahap berikutnya.

C. Pelabelan Data

Data hasil dari *preprocessing* akan dilakukan pelabelan dengan *Inset Lexicon* untuk menentukan apakah data tersebut tergolong sentimen negatif, positif, atau netral.

```
Jumlah Positif: 886
Jumlah Netral: 353
Jumlah Negatif: 2906
Total Data: 4145
```

	hasil_stemming	total_score	sentimen
0	pasang capres solusi pancasila ihwal rohingya	3.0	positif
1	gantung sikon biasa dampak positif lingkung be...	-17.0	negatif
2	dukung palestina usir rohingya dasar fasis mun...	0.0	netral
3	gampang hasut hoax rohingya	-4.0	negatif
4	mending tampung rohingya tempat duit pakai pak...	-4.0	negatif



Gambar 15. Hasil Pelabelan Data
Sumber: diolah dari data primer

Gambar 15 menunjukan jumlah data per label yang dibagi menjadi data negatif, positif, dan netral. Dari proses ini didapatkan 2906 tweet yang bernilai sentimen negatif, 886 positif, dan 353 netral.

D. Pembobotan Dsta

Data yang sudah melalui proses *preprocessing* dan pelabelan akan dihitung nilainya dengan *Term Frequency-Inverse Document Frequency* (TF-IDF).

	abad	abah	abai	abal	abalabal	abang	abar	abdel	abdul	abdur	...
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
...	...	...	...	...	...	...	...	...	...	...	...
3787	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
3788	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
3789	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
3790	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
3791	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...

	zioinist	zion	zionis	zionisme	zionist	zionistnya	zionists	zodiak
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...
3787	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3788	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3789	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3790	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3791	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Gambar 16. Hasil Pembobotan Data (TF-IDF)
Sumber: diolah dari data primer

Gambar 16 adalah hasil dari pembobotan untuk masing-masing kata dari dataset tweet. Adapun berikut ini contoh hasil pembobotan data yang memiliki nilai bobot selain 0.

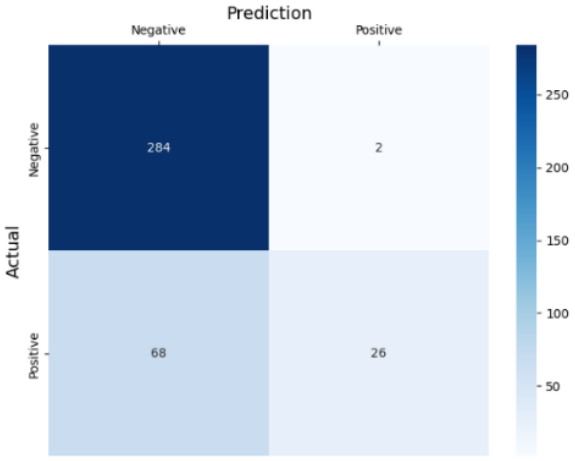
	text	Term	TF-IDF
0	0	capres	0.356894
1	0	ihwal	0.517947
2	0	pancasila	0.503232
3	0	pasang	0.472240
4	0	rohingya	0.068008
...	...	...	...
46919	3791	rohingya	0.034092
46920	3791	rumah	0.147212
46921	3791	sadar	0.183528
46922	3791	tinggal	0.147212
46923	3791	usir	0.115101

Gambar 17. Hasil TF-IDF Pelabelan InSet Lexicon
Sumber: diolah dari data primer

E. Klasifikasi SVM dan Evaluasi Model

Data yang sudah berhasil dikumpulkan, disesuaikan formatnya dengan kebutuhan sistem, diberi label, dan dihitung bobotnya akan dilakukan klasifikasi dengan *Support Vector Machine* (SVM) dengan beberapa parameter untuk menguji kemampuan model dalam berbagai kondisi. Terdapat sembilan skenario pengujian dengan parameter 0.1, 1, dan 10 serta pembagian data 90% data latih 10% data uji, 80% data latih 20% data uji, dan 70% data latih 30% data uji.

1. Skenario 1 (C = 0.1, 90% data latih 10% data uji)



Gambar 18. Confussion Matrix Skenario 1
Sumber: diolah dari data primer

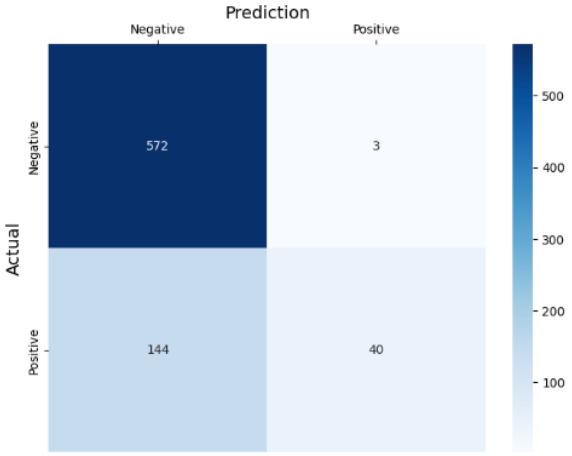
Hasil *confusion matrix* pada gambar 17, menunjukan 310 dari 380 data berhasil diklasifikasikan dengan benar oleh model.

```
Accuracy: 0.8157894736842105
Classification Report:
              precision    recall  f1-score   support

     negatif       0.81      0.99      0.89       286
     positif       0.93      0.28      0.43        94

    accuracy                           0.82       380
   macro avg       0.87      0.63      0.66       380
weighted avg       0.84      0.82      0.78       380
```

Gambar 19. Classification Report Skenario 1
Sumber: diolah dari data primer

Model mencapai akurasi 81,5%, *precision* 93% untuk data positif dan 81% untuk data negatif. Dari 380 data, 26 data sentimen positif dan 284 data sentimen negatif berhasil terklasifikasi dengan benar. Nilai *recall* yang didapatkan untuk sentimen positif adalah 28% dan 99% untuk sentimen negatif, serta *f1-score* 43% untuk sentimen positif dan 89% untuk sentimen negatif.

2. Skenario 2 (C = 0.1, 80% data latih 20% data uji)



Gambar 20. Confussion Matrix Skenario 2
Sumber: diolah dari data primer

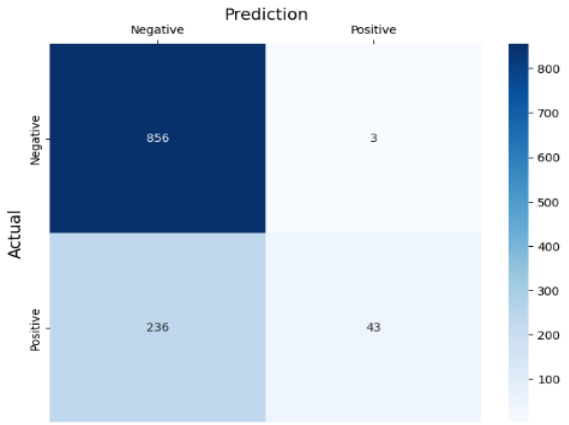
Hasil *confusion matrix* pada gambar 18, menunjukan 612 dari 759 data berhasil diklasifikasikan dengan benar oleh model.

```
Accuracy: 0.8063241106719368
Classification Report:
              precision    recall  f1-score   support

     negatif       0.80      0.99      0.89       575
     positif       0.93      0.22      0.35       184

    accuracy                           0.81       759
   macro avg       0.86      0.61      0.62       759
weighted avg       0.83      0.81      0.76       759
```

Gambar 21. Classification Report Skenario 2
Sumber: diolah dari data primer

Model mencapai akurasi 80,6%, *precision* 93% untuk data positif dan 80% untuk data negatif. Dari 759 data, 40 data sentimen positif dan 572 data sentimen negatif berhasil terklasifikasi dengan benar. Nilai *recall* yang didapatkan untuk sentimen positif adalah 22% dan 99% untuk sentimen negatif, serta *f1-score* 35% untuk sentimen positif dan 89% untuk sentimen negatif.

3. Skenario 3 (C = 0.1, 70% data latih 30% data uji)



Gambar 22. Confussion Matrix Skenario 3
Sumber: diolah dari data primer

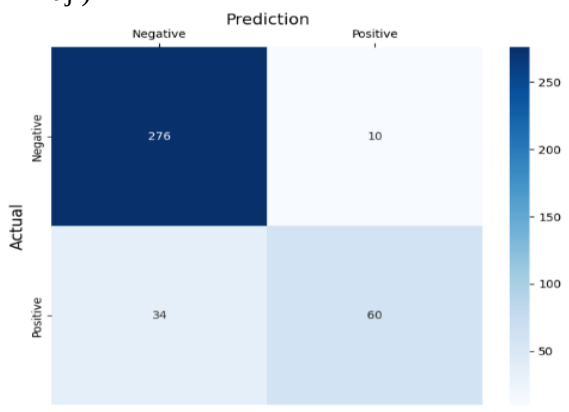
Hasil *confusion matrix* pada gambar 19, menunjukan 899 dari 1138 data berhasil diklasifikasikan dengan benar oleh model**.**

```
Accuracy: 0.7899824253075571
Classification Report:
              precision    recall  f1-score   support

     negatif       0.78      1.00      0.88       859
     positif       0.93      0.15      0.26       279

    accuracy                           0.79      1138
   macro avg       0.86      0.58      0.57      1138
weighted avg       0.82      0.79      0.73      1138
```

Gambar 23. Classification Report Skenario 3
Sumber: diolah dari data primer

Model mencapai akurasi 78,9%, *precision* 93% untuk data positif dan 78% untuk data negatif. Dari 1138 data, 43 data sentimen positif dan 856 data sentimen negatif berhasil terklasifikasi dengan benar. Nilai *recall* yang didapatkan untuk sentimen positif adalah 15% dan 100% untuk sentimen negatif, serta *f1-score* 26% untuk sentimen positif dan 88% untuk sentimen negatif.

4. Skenario 4 (C = 1, 90% data latih 10% data uji)



Gambar 24. Confussion Matrix Skenario 4
Sumber: diolah dari data primer

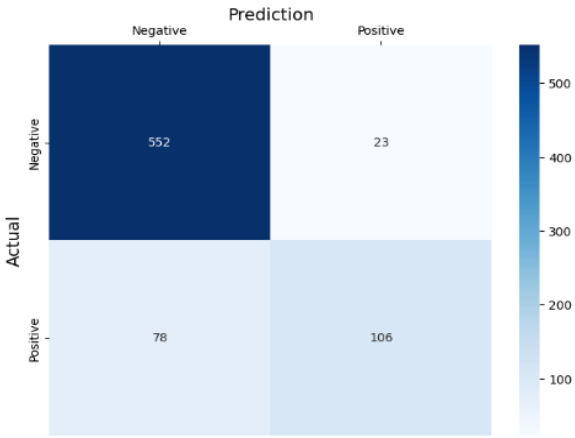
Hasil *confusion matrix* pada gambar 20, menunjukan 336 dari 380 data berhasil diklasifikasikan dengan benar oleh model.

```
Accuracy: 0.8842105263157894
Classification Report:
              precision    recall  f1-score   support

     negatif       0.89      0.97      0.93       286
     positif       0.86      0.64      0.73        94

    accuracy                           0.88       380
   macro avg       0.87      0.80      0.83       380
weighted avg       0.88      0.88      0.88       380
```

Gambar 25. Classification Report Skenario 4
Sumber: diolah dari data primer

Model mencapai akurasi 88,4%, *precision* 86% untuk data positif dan 89% untuk data negatif. Dari 380 data, 60 data sentimen positif dan 276 data sentimen negatif berhasil terklasifikasi dengan benar. Nilai *recall* yang didapatkan untuk sentimen positif adalah 64% dan 97% untuk sentimen negatif, serta *f1-score* 73% untuk sentimen positif dan 93% untuk sentimen negatif.

5. Skenario 2 (C = 1, 80% data latih 20% data uji)



Gambar 26. Confussion Matrix Skenario 5
Sumber: diolah dari data primer

Hasil *confusion matrix* pada gambar 18, menunjukan 658 dari 759 data berhasil diklasifikasikan dengan benar oleh model**.**

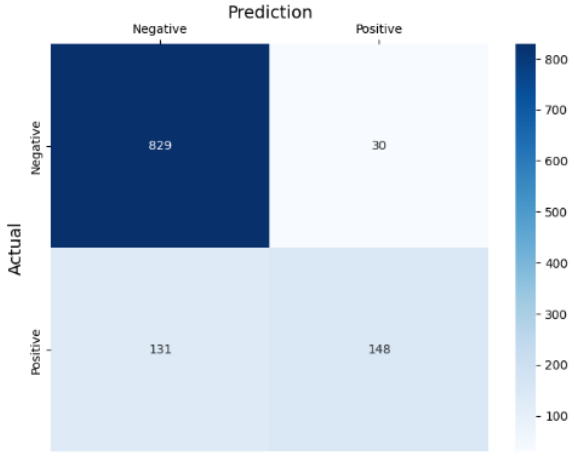
```
Accuracy: 0.8669301712779973
Classification Report:
              precision    recall  f1-score   support

     negatif       0.88      0.96      0.92       575
     positif       0.82      0.58      0.68       184

    accuracy                           0.87       759
   macro avg       0.85      0.77      0.80       759
weighted avg       0.86      0.87      0.86       759
```

Gambar 27. Classification Report Skenario 5
Sumber: diolah dari data primer

Model mencapai akurasi 86,6%, *precision* 82% untuk data positif dan 88% untuk data negatif. Dari 759 data, 106 data sentimen positif dan 552 data sentimen negatif berhasil terklasifikasi dengan benar. Nilai *recall* yang

didapatkan untuk sentimen positif adalah 58% dan 96% untuk sentimen negatif, serta *f1-score* 68% untuk sentimen positif dan 92% untuk sentimen negatif.

6. Skenario 3 (C = 1, 70% data latih 30% data uji)



Gambar 28. Confussion Matrix Skenario 6
Sumber: diolah dari data primer

Hasil *confusion matrix* pada gambar 19, menunjukan 977 dari 1138 data berhasil diklasifikasikan dengan benar oleh model**.**

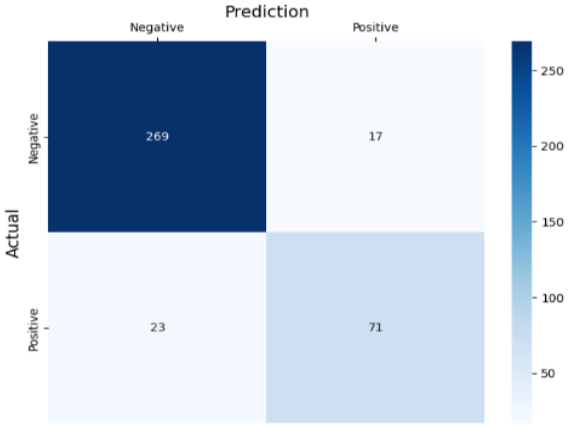
Accuracy: 0.8585237258347979
Classification Report:

	precision	recall	f1-score	support
negatif	0.86	0.97	0.91	859
positif	0.83	0.53	0.65	279
accuracy			0.86	1138
macro avg	0.85	0.75	0.78	1138
weighted avg	0.86	0.86	0.85	1138

Gambar 29. Classification Report Skenario 6
Sumber: diolah dari data primer

Model mencapai akurasi 85,8%, *precision* 83% untuk data positif dan 86% untuk data negatif. Dari 1138 data, 148 data sentimen positif dan 829 data sentimen negatif berhasil terklasifikasi dengan benar. Nilai *recall* yang didapatkan untuk sentimen positif adalah 53% dan 97% untuk sentimen negatif, serta *f1-score* 65% untuk sentimen positif dan 91% untuk sentimen negatif.

7. Skenario 7 (C = 10, 90% data latih 10% data uji)



Gambar 30. Confussion Matrix Skenario 7
Sumber: diolah dari data primer

Hasil *confusion matrix* pada gambar 20, menunjukan 340 dari 380 data berhasil diklasifikasikan dengan benar oleh model.

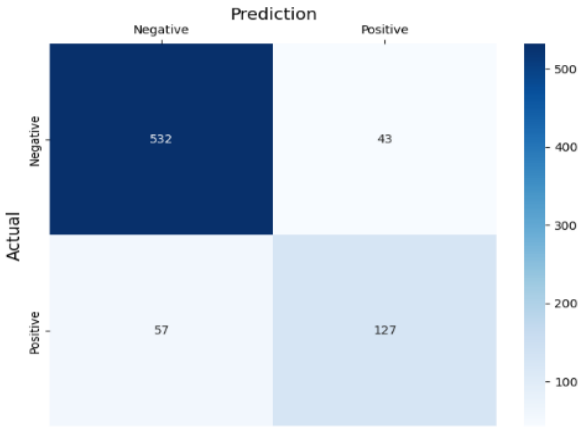
Accuracy: 0.8947368421052632
Classification Report:

	precision	recall	f1-score	support
negatif	0.92	0.94	0.93	286
positif	0.81	0.76	0.78	94
accuracy			0.89	380
macro avg	0.86	0.85	0.86	380
weighted avg	0.89	0.89	0.89	380

Gambar 31. Classification Report Skenario 7
Sumber: diolah dari data primer

Model mencapai akurasi 89,4%, *precision* 81% untuk data positif dan 92% untuk data negatif. Dari 380 data, 71 data sentimen positif dan 269 data sentimen negatif berhasil terklasifikasi dengan benar. Nilai *recall* yang didapatkan untuk sentimen positif adalah 76% dan 94% untuk sentimen negatif, serta *f1-score* 78% untuk sentimen positif dan 93% untuk sentimen negatif.

8. Skenario 8 (C = 10, 80% data latih 20% data uji)



Gambar 32. Confussion Matrix Skenario 8
Sumber: diolah dari data primer

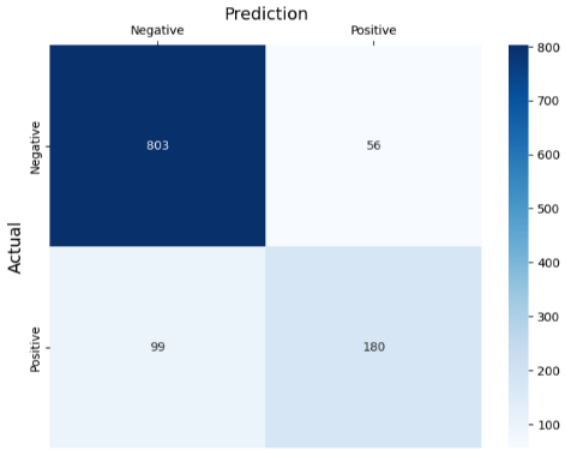
Hasil *confusion matrix* pada gambar 18, menunjukan 659 dari 759 data berhasil diklasifikasikan dengan benar oleh model**.**

```
Accuracy: 0.8682476943346509
Classification Report:
              precision    recall  f1-score   support

     negatif       0.90      0.93      0.91       575
     positif       0.75      0.69      0.72       184

    accuracy                           0.87       759
   macro avg       0.83      0.81      0.82       759
weighted avg       0.87      0.87      0.87       759
```

Gambar 33 Classification Report Skenario 8
Sumber: diolah dari data primer

Model mencapai akurasi 86,8%, *precision* 75% untuk data positif dan 90% untuk data negatif. Dari 759 data, 127 data sentimen positif dan 532 data sentimen negatif berhasil terklasifikasi dengan benar. Nilai *recall* yang didapatkan untuk sentimen positif adalah 69% dan 93% untuk sentimen negatif, serta *f1-score* 72% untuk sentimen positif dan 91% untuk sentimen negatif.

9. Skenario 9 (C = 10, 70% data latih 30% data uji)



Gambar 34. Confussion Matrix Skenario 9
Sumber: diolah dari data primer

Hasil *confusion matrix* pada gambar 19, menunjukan 983 dari 1138 data berhasil diklasifikasikan dengan benar oleh model**.**

```
Accuracy: 0.8637961335676626
Classification Report:
              precision    recall  f1-score   support

     negatif       0.89      0.93      0.91       859
     positif       0.76      0.65      0.70       279

    accuracy                           0.86      1138
   macro avg       0.83      0.79      0.81      1138
weighted avg       0.86      0.86      0.86      1138
```

Gambar 35. Classification Report Skenario 9
Sumber: diolah dari data primer

Model mencapai akurasi 86,3%, *precision* 76% untuk data positif dan 89% untuk data negatif. Dari 1138 data, hasil *precision* menunjukan 180 data sentimen positif dan 803 data sentimen negatif berhasil terklasifikasi dengan benar. Nilai *recall* yang didapatkan untuk sentimen positif adalah 65% dan 93% untuk sentimen negatif, serta *f1-score* 70% untuk sentimen positif dan 91% untuk sentimen negatif.

Setiap skenario pengujian memiliki nilai parameter dan hasil yang berbeda. Hal yang dapat dibandingkan adalah nilai akurasi dari semua hasil skenario pengujian. Berikut ini adalah rangkuman dari keseluruhan hasil skenario pengujian.

Tabel 1. Hasil Pengujian Skenario

Parameter C	Data Latih	Data Uji	Nilai Akurasi	Urutan Skenario
0.1	90%	10%	81.5%	Skenario 1
	80%	20%	80.6%	Skenario 2
	70%	30%	78.9%	Skenario 3
1	90%	10%	88.4%	Skenario 4
	80%	20%	86.6%	Skenario 5
	70%	30%	85.8%	Skenario 6
10	90%	10%	89.4%	Skenario 7
	80%	20%	86.8%	Skenario 8
	70%	30%	86.3%	Skenario 9

Sumber: diolah dari data primer

Rata-rata nilai akurasi dari keseluruhan skenario pengujian adalah 84,92% dengan nilai akurasi terendah yaitu 78,9% dengan parameter 0,1 serta 70% data uji dan 30% data latih. Sedangkan nilai akurasi tertinggi yaitu 89,4% dengan parameter 10 serta pembagian data 90% data latih dan 10% data uji.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, hasil dari proses pelabelan data menggunakan InSet Lexicon menghasilkan 886 sentimen positif, 2906 sentimen negatif, dan 353 sentimen netral. Hasil akurasi tertinggi yaitu 89.4% didapatkan oleh skenario pengujian dengan parameter C = 10 serta pembagian data 90% untuk data latih dan 10% untuk data uji. Sedangkan, hasil akurasi terendah yaitu 78.9%

didapatkan oleh skenario pengujian dengan parameter C = 0.1 serta pembagian data 90% untuk data latih dan 10% untuk data uji. Nilai rata-rata yang dihasilkan dari seluruh skenario pengujian yang telah dilakukan yaitu 89.4%. Saran untuk penelitian selanjutnya menggunakan algoritma klasikasi yang berbeda seperti Naïve Bayes serta menggunakan metode pelabelan data yang berbeda juga seperti VADER Lexicon.

DAFTAR PUSTAKA

Abdillah, T., Khaira, U., & Hutabarat, B. F. (2024). Komparasi Metode Naive Bayes dan K-Nearest Neighbors Terhadap Analisis Sentimen Pengguna Aplikasi Zenius. *Jurnal PROCESSOR*, *19*(1). https://doi.org/10.33998/processor.2024.19.1.1596

Berlianti, P. M., & Hidayat, E. Y. (2024). Implementasi Naïve Bayes Classifier untuk Sentimen Produk Kecantikan Berdasarkan Ulasan Female Daily. *The Indonesian Journal of Computer Science*, *13*(6), 10248–10256. https://doi.org/10.33022/ijcs.v13i6.4499

Dikananda, F., Rinaldi Dikananda, A., & Anwar, S. (2024). Analisis Sentimen Terhadap Marketplace Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(4), 8219–8225. https://doi.org/10.36040/jati.v8i4.10965

Fahrezi, A., & Solichin, A. (2024). *Analisis Sentimen Ulasan Aplikasi Mobile Jkn Pada Play Sentiment Analysis of the Jkn Mobile Application Reviews on the Play Store Using the Multinomial Naïve*. *3*(September), 1–10.

Hasbi, B. I., & Putro, I. S. (2025). Perbandingan Kinerja Algoritma Naive Bayes Dan Support Vector Machine untuk Analisis Sentimen Ulasan Produk E- Commerce Menggunakan Pembobotan TF-IDF. *Jurnal Komputer Dan Informatika*, *9*(2), 123–133. https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/4949

Hasibuan, H., Andrianto, R., Budiarti, P., Putri, R., Informasi, S., Teknologi, I., & Lawas, P. (2024). Analisis Sentimen Kepuasan Konsumen E-Commerce Tiktok Shop Menggunakan Metode Naive Bayes. *Jurnal Pendidikan Tambusai*, *8*(2), 33941–33950. https://jptam.org/index.php/jptam/article/view/18809

Kharde, V. A., & Sonawane, S. S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. International Journal of Computer Applications, 139(11), 975–8887. http://ai.stanford.

Krisdiyanto, T. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi, 7(1), 32. https://doi.org/10.24014/coreit.v7i1.12945

Muhammad, N. (2023, November 14). Ini Media Sosial yang Dipakai Anak Muda untuk Akses Informasi Politik. Databoks. https://databoks.katadata.co.id/politik/statistik/7374885aba94635/ini-media-sosial-yang-dipakai-anak-muda-untuk-akses-informasi-politik

Mulyadi, A. H., & Lestari, S. (2022). Analisis Sentimen Terhadap Sekolah Saat Covid-19 Pada Twitter Menggunakan Metode Lexicon Based. Jurnal Informatika Dan Teknologi Komputer, 03(01), 17–23. https://ejurnalunsam.id/index.php/jicom/

Musfiroh, D., Khaira, U., Utomo, P. E. P., & Suratno, T. (2021). Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 1, 24–33.

Muslim, S. N. S., Nurdiyansyah, F., & Rahman, A. Y. (2024). Perbandingan Algoritma Naive Bayes Dan KNN Dalam Analisis Sentimen Ulasan Pengguna Aplikasi Capcut. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3S1), 3588–3596. https://doi.org/10.23960/jitet.v12i3S1.5156

Ni'mah, A. T., & Arifin, A. Z. (2020). Perbandingan Metode Term Weighting terhadap Hasil Klasifikasi Teks pada Dataset Terjemahan Kitab Hadis. *Rekayasa*, *13*(2), 172–180. https://doi.org/10.21107/rekayasa.v13i2.6412

Prasetya, Y. N., Winarso, D., & Syahril. (2021). Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-

19. JURNAL FASILKOM, 11(2), 97–103.

Pratmanto, D., Imaniawan, F. F. D., & Maarif, V. (2023). Analisis Sentimen Pada Ulasan Pengguna Aplikasi Identitas Kependudukan Digital Dengan Metode Naive Bayes Dan K-Nearest. *Computatio : Journal of Computer Science and Information Systems*, *7*(2), 155–166. https://doi.org/10.24912/computatio.v7i2.26322

Pratmanto, D., Widayanto, A., Meisella Kristania, Y., & Wijianto, R. (2024). Analisis Perbandingan Algoritma Naive Bayes Dan KNN Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Vidio Di Google Play Store. *CONTEN : Computer and Network Technology*, *4*(2), 119–124. http://jurnal.bsi.ac.id/index.php/conten

R. Baidi. (2024, January 19). Kewajiban Kemanusiaan terhadap Pengungsi Rohingya. Detik.Com. https://news.detik.com/kolom/d-7146084/kewajiban-kemanusiaan-terhadap-pengungsi-rohingya.

Rayhan, M. R. E., Rudiman, R., & Fendy, F. Y. (2024). Perbandingan Metode K–Nearest Neighbor (KNN) Dan Naive Bayes Terhadap Analisis Sentimen Pada Pengguna E-Wallet Aplikasi Dana Menggunakan Fitur Ekstraksi TF-IDF. *Jurnal Teknologi Informasi: Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika*, *18*(2), 139–159. https://doi.org/10.47111/jti.v18i2.15009

Riki, Eliesse Dameria, S., Hermawan, A., Junaedi, & Kurnia, Y. (2025). Optimizing Sentiment Classification of E-Commerce Product Reviews: A Comparative Study of Naïve Bayes and SVM with SMO. *JUITA: Jurnal Informatika*, *13*(3), 287–295. https://doi.org/10.30595/juita.v13i3.26642

Roiqoh, S., Zaman, B., & Kartono, K. (2023). Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Mobile JKN dengan Lexicon Based dan Naïve Bayes. *Jurnal Media Informatika Budidarma*, *7*(3), 1582. https://doi.org/10.30865/mib.v7i3.6194

Rosid, M. A., Fitrani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020). Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi. *IOP Conference Series: Materials Science and Engineering*, *874*(1), 012017. https://doi.org/10.1088/1757-899X/874/1/012017

Styawati, Andi Nurkholis, Zaenal Abidin, & Heni Sulistiani. (2021). Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 5(5), 904–910. https://doi.org/10.29207/resti.v5i5.3380

Wiguna, D. P., & Ginting, S. E. (2026). Analisis Sentimen E-Commerce di Indonesia dengan Algoritma Naive Bayes (Studi Kasus Pada Platform Shopee, Tokopedia, Bukalapak, dan Lazada). *AICOM: Artificial Intelligence and Computing*, *01*(04), 109–114.

Wulandari, A. (2020). Aplikasi Support Vector Machine (SVM) Untuk Pencarian Binding Site Protein-Ligan. Jurnal Ilmiah Matematika, 8(2).

Yusuf, M., Gaja, R., Maulana, I., & Komarudin, O. (2023). Analisis Sentimen Opini Pengguna Aplikasi Vidio Pada Ulasan Playstore Menggunakan Algoritma Naive Bayes. JATI (Jurnal Mahasiswa Teknik Informatika), 7(4).

Zahra, N. Z., Farhanatussaidah, S., Afifah, N. N., Muthoharoh, L., Satria, A., & Manullang, M. C. T. (2026). Benchmarking Logistic Regression, SVM, Naive Bayes, and IndoBERT Fine-Tuning for Sentiment Analysis on Indonesian Product Reviews. *ArXiv Preprint ArXiv:2605.03439*.

