SKANIKA: Sistem Komputer dan Teknik Informatika
Volume 9, Nomor 2, Juli 2026, Halaman 258-269 E-ISSN: 2721-4788

Perbandingan Kinerja Hybrid InSet–SVM dan SentiStrength_id–SVM untuk Analisis Sentimen Platform Telemedicine Halodoc dan Alodokter

Asy Syifaur Roisah Rufaida
Fakultas Komunikasi dan Informatika, Sistem Informasi, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia
E-mail: [1*]asyifaur@ums.ac.id
(* : corresponding author)

Abstrak

Kualitas platform telemedicine dapat ditinjau melalui opini pengguna di media sosial menggunakan analisis sentimen. Penelitian analisis sentimen sebelumnya menunjukkan bahwa menggabungkan metode kamus dengan *machine learning* dapat meningkatkan akurasi analisis sentimen. Oleh karena itu penelitian ini akan membandingkan dua metode *hybrid* dengan kamus yang berbeda, yaitu *hybrid* Inset-SVM dan *hybrid* SentiStrength_id-SVM, untuk analisis sentimen terhadap platform *telemedicine* Halodoc dan Alodokter. Kedua metode dipilih karena memiliki mekanisme pembobotan sentimen yang sama, tetapi cakupan kosakata berbeda, Data diperoleh dari media sosial X (sebelumnya Twitter) menggunakan TWINT dengan kata kunci "Halodoc" dan "Alodokter" selama 1 Januari–31 Desember 2022. Data diberi label berdasarkan kedua kamus sentimen dan TF-IDF untuk kemudian dibagi menjadi data latih dan data uji untuk *machine learning* SVM (*Support Vector Machine*). Evaluasi model dilakukan dengan prosedur *3-repeated* 10-*fold cross validation*. Hasil perbandingan menunjukkan bahwa InSet-SVM lebih unggul dibandingkan dengan metode SentiStrength_id-SVM pada seluruh metrik evaluasi di kedua dataset. Pada dataset Halodoc, InSet-SVM menghasilkan nilai *precision* 85,92%, *recall* 86,24%, *f1-score* 85,76%, dan *accuracy* sebesar 86,24%. Pada dataset Alodokter, metode tersebut menghasilkan nilai *precision* 83,86%, *recall* 84,28%, *f1-score* 83,59%, dan *accuracy* sebesar 84,28%. Hasil tersebut menunjukkan bahwa penggunaan InSet sebagai leksikon pelabel awal menghasilkan model SVM yang lebih konsisten pada kedua dataset.

Kata kunci: analisis sentiment, *Indonesia Sentiment Lexicon, SentiStrength_id, Support Vector Machine*

Abstract

The quality of telemedicine platforms can be assessed through sentiment analysis of user opinions on social media. This study compared two hybrid methods, InSet-SVM and SentiStrength_id-SVM, for sentiment analysis of the Halodoc and Alodokter platforms. Both lexicons were selected because they apply similar sentiment-weighting mechanisms but differ in vocabulary coverage. Data were collected from X (formerly Twitter) using TWINT with the keywords "Halodoc" and "Alodokter" from January 1 to December 31, 2022. The data were labeled using each sentiment lexicon, while TF-IDF was applied for feature weighting before classification using a Support Vector Machine (SVM). The models were evaluated using three-times-repeated 10-fold cross-validation. The results showed that InSet-SVM outperformed SentiStrength_id-SVM across all evaluation metrics for both datasets. For Halodoc, InSet-SVM achieved 85.92% precision, 86.24% recall, an F1-score of 85.76%, and 86.24% accuracy. For Alodokter, it achieved 83.86% precision, 84.28% recall, an F1-score of 83.59%, and 84.28% accuracy. These findings indicate that using InSet as the initial labeling lexicon produces a more consistent SVM model across both datasets.

***Keywords**: sentiment analysis, Indonesia Sentiment Lexicon, SentiStrength_id, Support Vector Machine*

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong transformasi pelayanan kesehatan melalui telemedicine, yaitu penyelenggaraan layanan kesehatan jarak jauh dengan dukungan teknologi informasi dan komunikasi. Pandemi COVID-19 mempercepat adopsi telemedicine dan menjadikannya salah satu bentuk pelayanan kesehatan yang semakin umum digunakan [1], [2]. Peningkatan penggunaan platform telemedicine menuntut penyedia layanan untuk terus mengevaluasi kualitas dan pengalaman penggunanya. Salah satu sumber informasi yang dapat dimanfaatkan dalam evaluasi tersebut adalah opini pengguna yang disampaikan melalui media sosial dan platform digital [3]. Platform X (dulu Twitter) adalah salah satu platform media sosial

yang sering digunakan warganet sebagai tempat beropini, baik opini yang positif maupun negatif. Opini positif dan negatif ini dapat diklasifikasikan dengan cara analisis sentimen.

Analisis sentimen merupakan proses mengekstraksi opini, sikap, atau perasaan yang terdapat dalam data tekstual. Secara umum, analisis sentimen dapat dilakukan menggunakan pendekatan berbasis leksikon, *machine learning*, *deep learning*, maupun kombinasi dari beberapa pendekatan tersebut [4], [5]. Pendekatan berbasis leksikon menentukan polaritas berdasarkan bobot sentimen kata-kata dalam suatu teks, sedangkan pendekatan *machine learning* membangun model klasifikasi berdasarkan pola yang dipelajari dari data berlabel [5].

Penelitian yang menganalisis sentimen terhadap aplikasi *telemedicine* di antaranya adalah [6], [7], [8], [9]. Penelitian [6] menganalisis ulasan pengguna Halodoc, Alodokter, KlikDokter, SehatQ, dan YesDok menggunakan BiLSTM dan pemodelan topik LDA. Model BiLSTM dan Word2Vec yang digunakan menghasilkan accuracy sebesar 95%. Kemudian penelitian [7] menganalisis cuitan di X tentang aplikasi Halodoc, Alodokter, dan Klikdokter menggunakan algoritma *Naïve Bayes*. Selanjutnya penelitian [8], [9] membandingkan beberapa metode *machine learning* untuk menganalisis ulasan aplikasi Halodoc. Penelitian [8] membandingkan algoritma *K-Nearest Neighbour* (KNN), SVM dan *Naïve Bayes*, hasil menunjukkan bahwa algoritma dengan akurasi terbaik adalah KNN dengan akurasi sebesar 95%. Sementara penelitian [9] membandingkan algoritma *Naïve Bayes* dan SVM kernel RBF, hasil menunjukkan bahwa nilai akurasi yang lebih baik dihasilkan oleh *Naïve Bayes*, yaitu 87,77%.

Penelitian terbaru menunjukkan bahwa SVM memberikan kinerja yang baik dalam analisis sentimen pada berbagai domain. Penelitian [10] membandingkan algoritma Naive Bayes dan SVM untuk menganalisis sentimen pengguna True Wireless Stereo di media sosial X. Hasil penelitian menunjukkan bahwa SVM memperoleh akurasi sebesar 80,46% dan mengungguli Naive bAyes yang memperoleh akurasi sebesar 78,00%. Sementara itu [11] menerapkan TF-IDF dan beberapa jenis algoritma SVM pada ulasan Google Maps. SSVM LInearSSVC menghasilkan kinerja terbaik dengan akurasi sebesar 84%. Kedua penelitian tersebut menunjukkan bahwa SVM dapat digunakan secara efektif untuk klasifikasi sentimen, meskipun diterapkan pada objek dan sumber data yang berbeda.

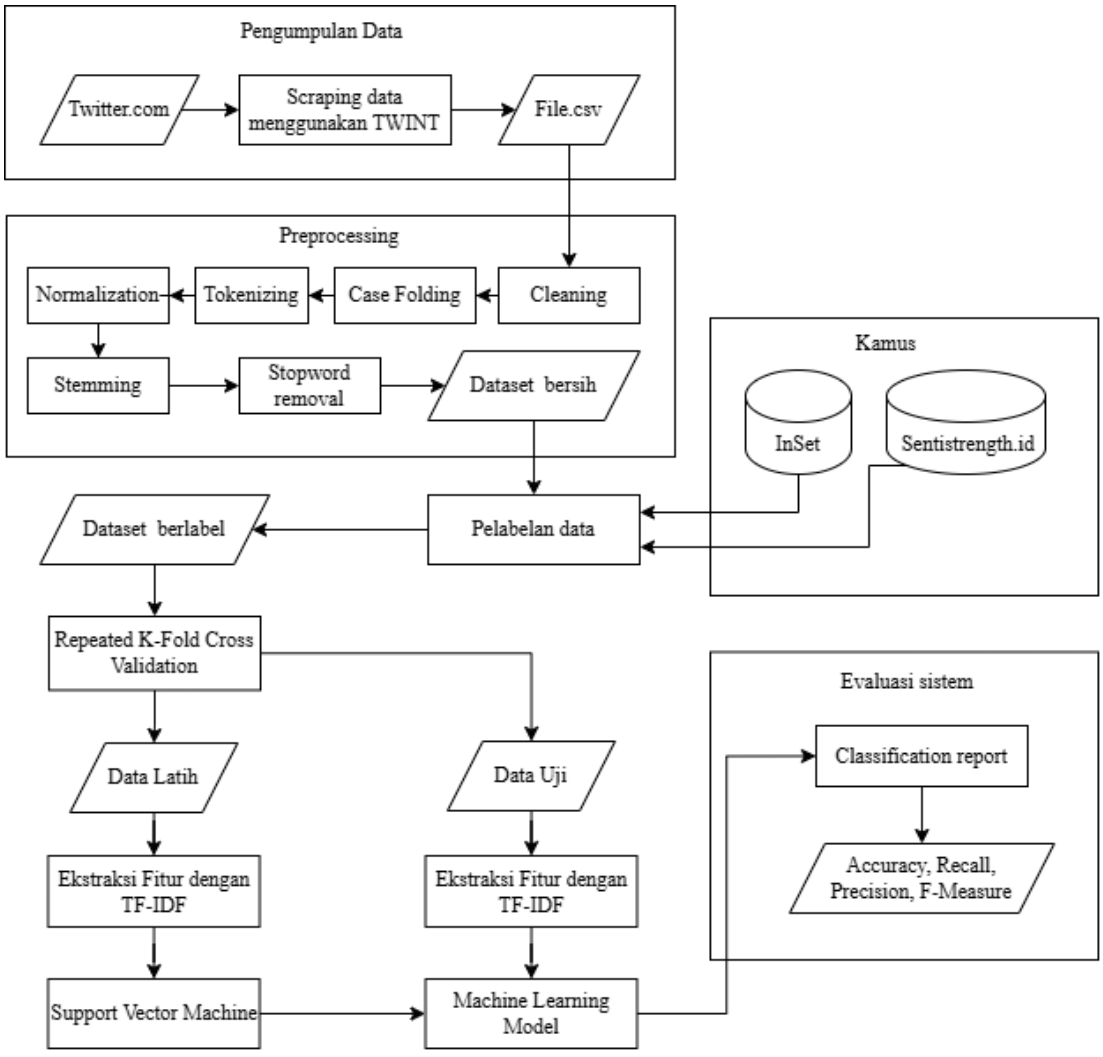
Terdapat beberapa penelitian yang bertujuan meningkatkan akurasi analisis sentimen dengan menggabungkan metode kamus dengan *machine learning*. Penelitian [4] membandingkan metode *hybrid* kamus-*machine learning*, metode *standalone machine learning*, dan metode kamus saja. Kamus yang digunakan dalam metode *hybrid* adalah SentiWordNet dan *machine learning* yang digunakan adalah SVM. Metode *hybrid* kamus-*machine learning* ini memiliki kinerja paling unggul (lebih dari 85%) dibandingkan metode *standalone machine learning* (seperti *Naïve Bayes*, SVM, KNN, *Random Forest,* dan *MaxEnt*) maupun metode kamus (seperti *Vader* dan *TextBlob*). Penelitian [12] membandingkan kinerja kamus InSet dan SentiStrength_id yang digabungkan dengan SVM untuk menganalisis sentimen pada domain kebijakan pemerintah khususnya dalam penanganan COVID-19. Hasil pengujian penelitian ini menunjukkan bahwa metode *hybrid* Inset-SVM menghasilkan akurasi yang paling unggul yaitu 83%. Sedangkan metode *hybrid* SentiStrength_id-SVM dan metode *standalone* SVM menghasilkan nilai akurasi yang sama yaitu 82%.

Berdasarkan penelitian-penelitian tersebut, disimpulkan bahwa kinerja analisis sentimen berhasil ditingkatkan dengan penggabungan kedua metode (metode kamus dan *machine learning*). Penelitian sebelumnya menunjukkan keunggulan InSet–SVM pada domain kebijakan penanganan COVID-19, tetapi hasil tersebut belum dapat langsung digeneralisasi ke domain telemedicine karena perbedaan konteks, istilah kesehatan, dan gaya bahasa pengguna. Penelitian telemedicine terdahulu juga lebih banyak membandingkan algoritma machine learning atau menggunakan satu sumber label, sehingga belum menjelaskan pengaruh pemilihan leksikon Indonesia terhadap kinerja classifier pada data X. Oleh karena itu penelitian ini akan membandingkan dua metode *hybrid* dengan kamus yang berbeda, yaitu *hybrid* Inset-SVM dan *hybrid* SentiStrength_id-SVM, untuk analisis sentimen terhadap platform *telemedicine* khususnya Halodoc dan Alodokter. Pemilihan metode ini disesuaikan berdasarkan penelitian [12] yang membandingkan kedua metode *hybrid* ini pada objek penelitian yang berbeda. Penelitian ini

dilakukan untuk mengetahui sentimen publik di media sosial X terhadap aplikasi Halodoc dan Alodokter, serta membandingkan kinerja kedua metode *hybrid* ini pada data sentimen tentang Halodoc dan Alodokter.

2. METODE PENELITIAN

Penelitian ini membandingkan dua kamus sentimen pada metode *hybrid* kamus-SVM. Kamus sentimen yang digunakan adalah kamus InSet [13] dan kamus SentiStrength_id [14]. Tahap penelitian yang dibutuhkan dalam mencapai hasil klasifikasi yang optimum, diantaranya adalah *preprocessing*, pelabelan data, ekstraksi fitur, pengembangan model klasifikasi, dan evaluasi sistem. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Data diambil dari X dengan cara *scrapping* menggunakan *library* Python yaitu TWINT (*Twitter Intelligent Tool*). Kata kunci yang digunakan adalah "Halodoc" dan "Alodokter". Data yang diambil hanyalah cuitan berbahasa Indonesia yang di-*posting* di antara tanggal 1 Januari 2022 hingga tanggal 31 Desember 2022. Cuitan dari akun X resmi Halodoc (@HalodocID) dan Alodokter (@alodokter) tidak diambil untuk menghindari bias. Data ini selanjutnya disimpan dalam bentuk file .csv (*Comma Separated Value*). Contoh data hasil *scrapping* X tersedia pada Tabel 1.

Tabel 1. Contoh Data Hasil Scrapping X

Cuitan
@adarazcl @schfess mjb kak, kalo mau konsul ke psikolog di halodoc tuh bayar apa ngga yaa?
@ohmybeautybank eh kok gelembung gt ya, ada cairannya kah nder? mending coba konsul sm dokter kulit di halodoc
@ilovefrapucino @ohmybeautybank bener, chat halodoc aja der pasti ada jawaban dari dokter kandungan langsung, kan sama2 chat di sosmed juga. cari yg paling murah aja tarifnya. semangat der

2.2. *Preprocessing*

Tahap *preprocessing* dilakukan untuk membersihkan dan menormalisasi teks sebelum digunakan dalam analisis sentimen. Tahap ini penting terutama pada data media sosial karena teks sering mengandung singkatan, kesalahan ejaan, bahasa tidak baku, tanda baca, dan unsur lain yang dapat menjadi *noise* [15]. Tahapan *preprocessing* dalam penelitian ini terdiri dari:

a. Cleaning
   Tahap ini menghapus elemen-elemen data yang tidak penting atau tidak mempengaruhi proses analisis. Elemen-elemen tersebut seperti angka, tanda baca, jarak kata yang berlebih dan lain-lain.
b. Case folding
   Pada tahap ini, semua huruf kapital pada kata diubah menjadi huruf kecil (*lowercase*).
c. Tokenizing
   Tahap ini memecah kalimat menjadi beberapa kata tunggal untuk mempermudah proses berikutnya.
d. Normalization
   Tahap ini memperbaiki kata yang tidak sesuai dengan ejaan sebenarnya, seperti salah ketik, singkatan dan kata tidak baku.
e. Stemming
   Tahap ini membuang imbuhan dari tiap kata untuk mengubahnya menjadi kata dasar.
f. Stopword removal
   Pada tahap ini, kata-kata yang dianggap tidak penting, seperti kata awalan maupun kata-kata yang tidak memberikan pengaruh terhadap proses analisis sentimen, dihilangkan.

Contoh data di tahap *preprocessing* tersedia pada tabel 2.

Tabel 2. Contoh *Preprocessing* Data

Tahap	Hasil
Data mentah	@18fesss Wkwkk, 500k cuma 5mnt pula.. mendingan ke Halodoc dl
Cleaning	Wkwkk, cuma pula.. mendingan ke Halodoc dl
Case folding	wkwkk, cuma pula.. mendingan ke halodoc dl
Tokenizing	['wkwkk', 'cuma', 'pula', 'mendingan', 'ke', 'halodoc', 'dl']
Normalization	['wkwkk', 'cuma', 'pula', 'mendingan', 'ke', 'halodoc', 'dulu']
Stemming	['wkwkk', 'cuma', 'pula', 'mending', 'ke', 'halodoc', 'dulu']
Stopword removal	['wkwkk', 'mending', 'halodoc']

2.3. Pelabelan Data

Data mentah yang sudah melalui tahap *preprocessing* selanjutnya dilabeli berdasarkan kamus InSet [13] dan SentiStrength_id [14]. Tiap kata dalam cuitan akan diberi bobot dan dihitung nilai sentimennya berdasarkan nilai polaritas yang ada pada kamus sentimen. Nilai sentimen akhir sebuah kalimat ditetapkan berdasarkan nilai positif terbesar dan nilai negatif terbesar dari seluruh kata penyusunnya. Sebelum dilabeli, tiap kalimat diberi nilai bawaan positif terbesar = 1 dan negatif terbesar = -1. Selanjutnya label sentimen akhir akan mengikuti persamaan (1).

$$\begin{aligned} if\ \max positive\ value > |\max negative\ value| &\rightarrow Positive \\ if\ \max positive\ value = |\max negative\ value| &\rightarrow Neutral \\ if\ \max positive\ value < |\max negative\ value| &\rightarrow Negative \end{aligned} \quad (1)$$

Contoh hasil pelabelan data tersedia pada Tabel 3.

Tabel 3. Contoh Pelabelan Data

Pelabelan Nilai Polaritas	Max Positive	Max Negative	Label Sentimen
['di halodoc tuh [-2] suka [4] diskon']	4	-2	Positif
['kalau mau ribet bisa pakai bpjs di halodoc pernah coba tapi banyak tidak cocok [-3] ']	1	-3	Negatif

2.4. Ekstraksi Fitur

Ekstraksi fitur adalah proses mengubah data berbentuk teks menjadi data numerik dan himpunan fitur dalam bentuk vektor yang kemudian digunakan dalam proses klasifikasi. Penelitian ini menggunakan fitur set TF-IDF Unigram. TF-IDF memberikan bobot pada suatu kata berdasarkan frekuensi kemunculannya dalam sebuah dokumen dan tingkat kelangkaannya pada keseluruhan dokumen. Kombinasi TF-IDF dan SVM telah banyak digunakan untuk merepresentasikan dan mengklasifikasikan data sentimen berbentuk teks [16], [17].

Perhitungan TF-IDF dilakukan melalui dua tahap, yaitu TF (*Term Frequency*) yang digunakan untuk menghitung banyak kata muncul pada suatu dokumen, dan IDF (*Inverse Document Frequency*) yang digunakan untuk menghitung bagaimana kata tersebut tersebar dalam suatu dokumen. Penelitian ini menghitung TF-IDF menggunakan *library* s*klearn.feature_extraction* di Python. Rumus perhitungan bobot TF-IDF yang digunakan pada *library sklearn* dapat dilihat pada persamaan (2).

$$W_{t,d} = tf_{t,d} \times ln\left(\frac{1+N}{1+df_t}\right) + 1 \quad (2)$$

Keterangan:
$W_{t,d}$: *Term Frequency - Inverse Document Frequency*
$tf_{t,d}$: Jumlah kata (*term*) t pada dokumen d
N : Jumlah dokumen teks
df_t : Jumlah dokumen teks yang mengandung kata (*term*) t

2.5. *Repeated* k-fold *cross validation*

Kinerja model dapat dievaluasi secara spesifik melalui prosedur *repeated k-fold cross validation*. *K-fold* adalah sebuah teknik pengukuran kinerja model dengan cara membagi dataset menjadi beberapa bagian (*fold*), di mana setiap bagian secara bergiliran digunakan sebagai data uji, sementara bagian lainnya sebagai data latih. Langkah ini dilakukan karena pembagian data yang berbeda dapat menghasilkan estimasi kinerja yang juga berbeda..

Penggunaan *repeated k-fold cross validation* ini mengulang prosedur *k-fold cross validation* sebanyak n kali. Pada setiap pengulangan, data akan diacak lagi sebelum dibagi menjadi lipatan (*folds*). Ini memungkinkan untuk mendapatkan variasi yang berbeda dalam pemisahan data pada setiap pengulangan. Penelitian ini menggunakan 10-*fold* dengan 3-*repeat* (pengulangan tiga kali), sehingga total *fold* adalah 30.

2.6. Pengembangan Model Klasifikasi

Data yang telah dilabeli menggunakan kamus sentimen dan diproses dengan TF-IDF selanjutnya dibagi menjadi data latih dan data uji dengan metode *k-fold cross validation*. Data latih berfungsi untuk melatih model *machine learning*, sedangkan data uji digunakan untuk menguji kinerja model dalam memprediksi label. Pada penelitian ini, algoritma *machine learning* yang digunakan adalah SVM (*Support Vector Machine*) [18].

2.7. Evaluasi Sistem

Tahap terakhir adalah evaluasi sistem yang bertujuan untuk menilai kinerja model klasifikasi yang telah dibangun. Metode evaluasi yang digunakan adalah *classification report* yang menunjukkan nilai *precision*, *recall*, dan *f1-score* untuk setiap label sentimen (positif maupun negatif) serta nilai *accuracy*. Perhitungan metrik evaluasi tersebut ditunjukkan pada persamaan (3-6).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (5)$$

$$Accuracy = \frac{TP + TN}{TN + TP + FP + FN} \quad (6)$$

Keterangan:
TP : Jumlah cuitan positif yang diprediksi benar oleh sistem
TN : Jumlah cuitan negatif yang diprediksi benar oleh sistem
FP : Jumlah cuitan negatif yang diprediksi sebagai cuitan positif oleh sistem
FN : Jumlah cuitan positif yang diprediksi sebagai cuitan negatif oleh sistem

Evaluasi dilakukan pada setiap *fold*, kemudian hasil dari seluruh *fold* dihitung nilai rata-ratanya untuk memperoleh gambaran kinerja model secara keseluruhan. Selanjutnya, hasil evaluasi dari masing-masing metode dibandingkan satu sama lain. Perbandingan tersebut digunakan untuk mengetahui metode *hybrid* dengan kamus sentiment mana yang memberikan performa terbaik.

3. HASIL DAN PEMBAHASAN

Hasil penelitian ini menyajikan hasil dari pengumpulan data hingga hasil evaluasi dari prediksi SVM dalam melakukan analisis sentimen dari cuitan tentang aplikasi Halodoc dan Alodokter.

3.1. Pengumpulan Data

Hasil *scrapping* data menggunakan TWINT dengan kata kunci "halodoc" pada rentang waktu dari 1 Januari 2022 hingga tanggal 31 Desember 2022 adalah 6543 cuitan. Sedangkan data dengan kata kunci "alodokter" pada rentang waktu yang sama adalah 2736 cuitan. Data kemudian disaring kembali dengan hanya mengambil cuitan berbahasa Indonesia. Setelah melalui penyaringan data, diperoleh 5768 cuitan tentang Halodoc dan 2519 cuitan tentang Alodokter. Kedua dataset diproses dan dievaluasi secara terpisah agar kinerja metode pada masing-masing platform dapat dibandingkan. Data ini terbagi dalam 12 bulan dengan distribusi yang ditunjukkan pada tabel 4. Tabel 4 menunjukkan bahwa proses penyaringan bahasa mengurangi data Halodoc dari 6.543 menjadi 5.768 cuitan dan data Alodokter dari 2.736 menjadi 2.519 cuitan. Dengan demikian, 775 cuitan Halodoc dan 217 cuitan Alodokter dikeluarkan karena tidak memenuhi kriteria bahasa Indonesia.

Tabel 4. Distribusi Data Cuitan

Bulan	Jumlah Cuitan Halodoc		Jumlah Cuitan Alodokter	
	Scrapping Awal	Berbahasa Indonesia	Scrapping Awal	Berbahasa Indonesia
1	513	460	205	185
2	536	491	203	195
3	525	485	168	159
4	554	497	229	213
5	611	559	253	233
6	539	494	243	219
7	556	480	269	250
8	478	415	195	177
9	782	636	216	206
10	514	444	281	265
11	463	409	260	234
12	472	398	214	183
Total	**6543**	**5768**	**2736**	**2519**

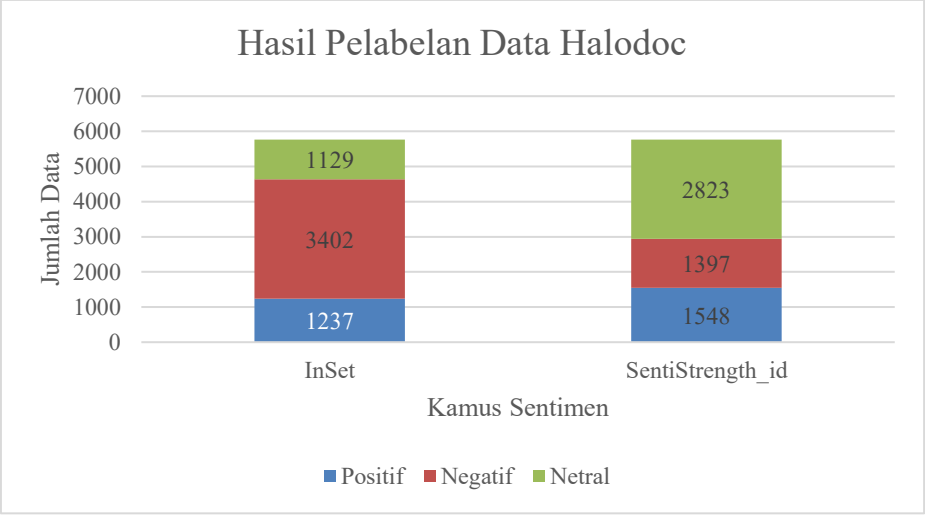
Data ini kemudian disimpan dalam bentuk file .csv. Salah satu data cuitan yang diperoleh dapat dilihat pada tabel 5.

Tabel 5. Sebagian Data Cuitan

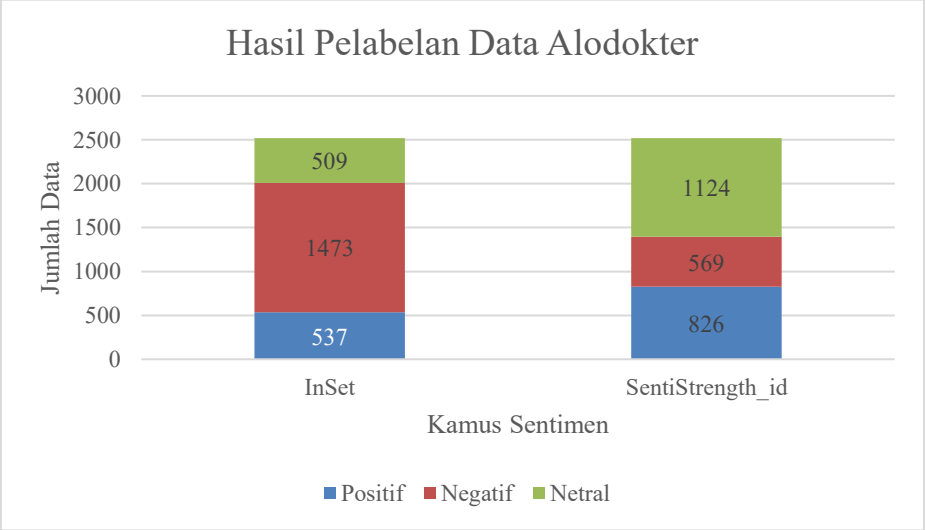
Cuitan Halodoc	@ZOO_FESS Tapi coba konsul ke halodoc aja dulu, besok baru dibawa ke vet. Semoga mengnya ga kenapaÂ²
Cuitan Alodokter	jam segini konsul alodokter nemu dokter baik banget mo nanges

3.2. Pelabelan Data

Data yang sudah melalui tahap *preprocessing* kemudian dilabeli berdasarkan kamus sentimen InSet dan juga kamus SentiStrength_id. Pelabelan data menggunakan kamus sentimen yang berbeda akan menghasilkan klasifikasi yang berbeda pula. Penggunaan kamus InSet untuk pelabelan data pada dokumen Halodoc menghasilkan 1237 label positif, 3402 label negatif dan 1129 label netral, serta pada dokumen Alodokter menghasilkan 537 label positif, 1473 label negatif dan 509 label netral. Sementara itu, penggunaan kamus SentiStrength_id untuk pelabelan data pada dokumen Halodoc menghasilkan 1548 label positif, 1397 label negatif dan 2823 label netral, serta pada dokumen Alodokter menghasilkan 826 label positif, 569 label negatif dan 1124 label netral. Distribusi hasil pelabelan data untuk tiap dokumen menggunakan kedua kamus berbeda ditunjukkan pada Gambar 2 dan 3.



Gambar 2. Distribusi Hasil Pelabelan Data Halodoc

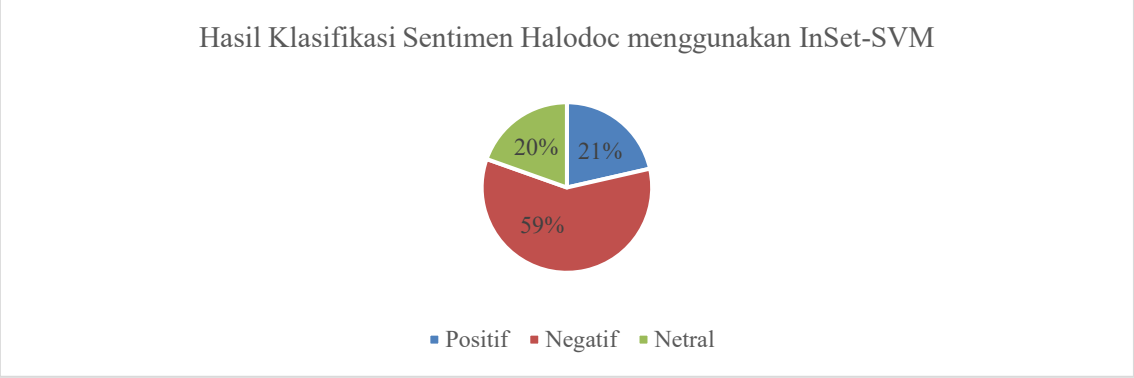


Gambar 3. Distribusi Hasil Pelabelan Data Alodokter

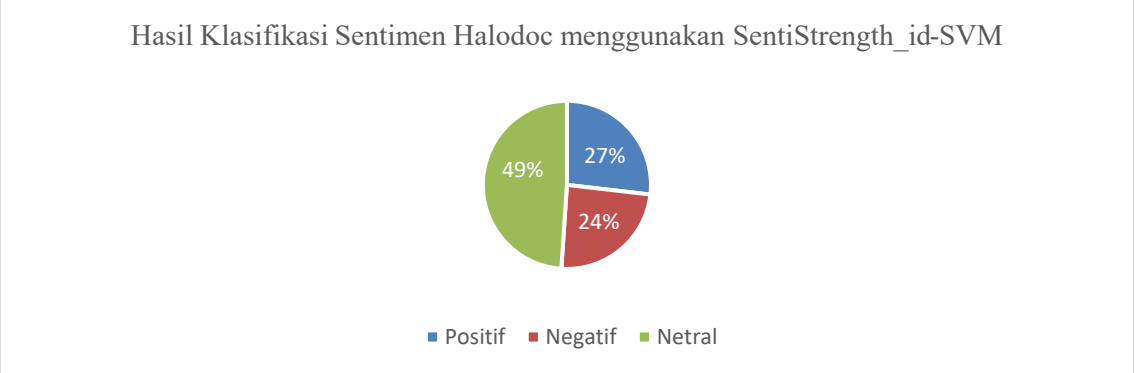
Sebagaimana yang ditunjukkan pada grafik di Gambar 2 dan 3, pelabelan data menggunakan kamus sentimen yang berbeda akan menghasilkan klasifikasi yang berbeda pula. Pada pelabelan data menggunakan kamus SentiStrength_id data paling banyak diberi label netral, sedangkan pada metode InSet data paling banyak diberi label negatif. Beberapa data dilabeli netral karena tidak ada kata yang diberi bobot (bobot=0), dengan kata lain kata-kata dalam kamus sentimen kurang beragam. Kondisi ini sesuai dengan kamus SentiStrength_id yang memiliki daftar kata lebih sedikit dibanding kamus InSet. Hasil evaluasi penelitian tersebut menunjukkan bahwa tiap kamus memiliki nilai kinerja yang berbeda, semakin bervariasi kata-kata dalam kamus sentimen maka akan meningkat juga nilai akurasinya.

3.3. Klasifikasi Sentimen

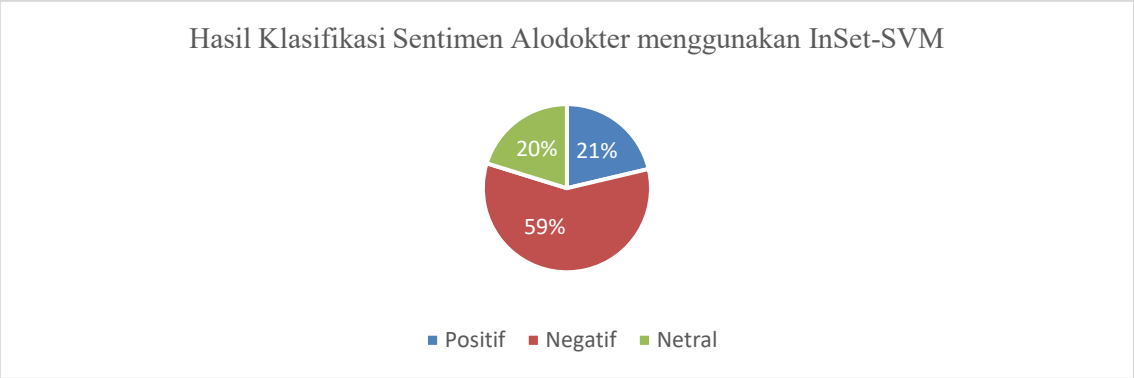
Gambar 4–7 memperlihatkan bahwa distribusi label sangat dipengaruhi oleh leksikon yang digunakan. Pada Halodoc, InSet menghasilkan 21% positif, 59% negatif, dan 20% netral (Gambar 4), sedangkan SentiStrength_id menghasilkan 27% positif, 24% negatif, dan 49% netral (Gambar 5). Pada Alodokter, InSet menghasilkan 21% positif, 59% negatif, dan 20% netral (Gambar 6), sedangkan SentiStrength_id menghasilkan 33% positif, 22% negatif, dan 45% netral (Gambar 7). Dominasi label netral pada SentiStrength_id menunjukkan lebih banyak cuitan memperoleh skor nol karena kata-katanya tidak terpetakan dalam leksikon. Sebaliknya, cakupan InSet yang lebih luas menyebabkan lebih banyak cuitan memperoleh polaritas



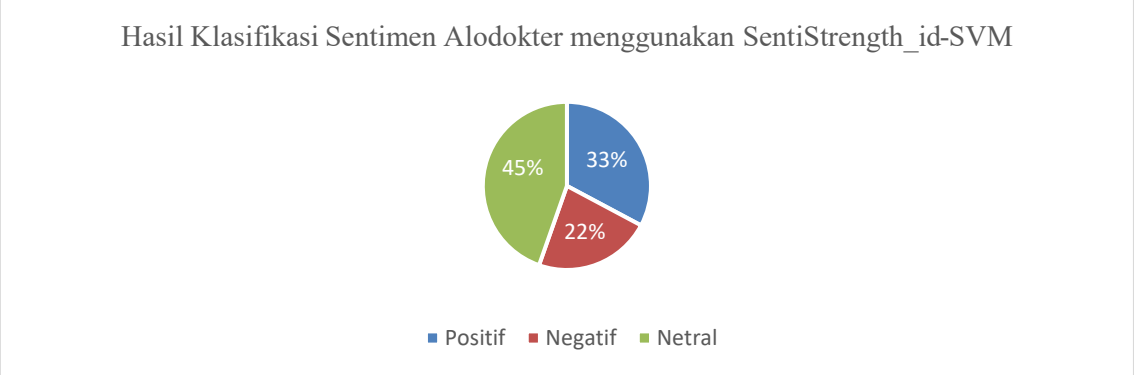
Gambar 4. Persentase Hasil Klasifikasi Sentimen Halodoc menggunakan Inset-SVM



Gambar 5. Persentase Hasil Klasifikasi Sentimen Halodoc menggunakan SentiStrength_id-SVM



Gambar 6. Persentase Hasil Klasifikasi Sentimen Alodokter menggunakan Inset-SVM



Gambar 7. Persentase Hasil Klasifikasi Sentimen Alodokter menggunakan SentiStrength_id-SVM

Berdasarkan hasil klasifikasi tersebut, dapat disimpulkan bahwa metode *hybrid* InSet-SVM cenderung menghasilkan paling banyak label negatif, sementara metode *hybrid* SentiStrength_id-SVM didominasi oleh hasil label netral. Sentimen netral tentang Halodoc mencakup pertanyaan tentang Halodoc, berita tentang Halodoc, dan cuitan tentang Halodoc yang tidak mengungkapkan pendapat tentang aplikasi tersebut. Sentimen positif tentang Halodoc mencakup saran untuk mencoba konsultasi di Halodoc serta biaya yang rendah dari layanan Halodoc. Sementara itu, sentimen negatif tentang Halodoc adalah harus membayar selama konsultasi di Halodoc, tidak ada perubahan setelah menerima perawatan dari Halodoc, tenaga medis yang dianggap tidak kompeten, dan pengiriman obat yang rumit.

Sentimen netral tentang Alodokter mencakup berita tentang Alodokter dan cuitan tentang Alodokter yang tidak mengungkapkan pendapat tentang aplikasi tersebut. Sentimen positif tentang Alodokter mencakup saran untuk mencoba konsultasi di Alodokter serta biaya yang rendah dari layanan Alodokter. Sementara itu, sentimen negatif tentang Alodokter termasuk artikel dengan sumber yang tidak jelas, tenaga medis yang dianggap tidak kompeten, dan letak apotek yang jauh.

3.4. Evaluasi Model

Evaluasi kinerja model secara spesifik dilakukan melalui prosedur *repeated k-fold cross validation*. Pada penelitian ini digunakan 10-*fold* dengan tiga kali pengulangan (3-*repeat*), sehingga secara keseluruhan terdapat 30 *fold*. Pada setiap *fold*, data latih dan data uji terlebih dahulu melalui proses ekstraksi fitur menggunakan TF-IDF sebelum digunakan untuk melatih model. Hasil evaluasi dari tiap *fold* kemudian dikumpulkan dan dihitung nilai rata-ratanya untuk memperoleh gambaran kinerja model secara menyeluruh. Metode evaluasi yang digunakan adalah *classification report* yang memuat nilai *precision*, *recall*, *f1-score* serta *accuracy*.

Pada penelitian ini terdapat ketidakseimbangan antara jumlah data kelas positif dan kelas negatif (*imbalance data*), sehingga mempertimbangkan nilai akurasi saja mungkin tidak memberikan gambaran yang akurat tentang kualitas model. Ketidakseimbangan kelas terbentuk

sejak tahap pelabelan berbasis leksikon. Pada InSet, bobot kata negatif yang terdeteksi menyebabkan kelas negatif menjadi mayoritas. Pada SentiStrength_id, cakupan kosakata yang lebih terbatas menyebabkan banyak cuitan tidak memperoleh skor polaritas dan dilabeli netral. Oleh karena itu ada beberapa nilai lain yang perlu dipertimbangkan lebih dalam yaitu: *precision*, *recall*, *f1-score*. Nilai *precision* yang lebih tinggi menunjukkan bahwa model lebih akurat dalam mengidentifikasi kelas positif. Nilai ini perlu dipertimbangkan jika fokus utama adalah mengidentifikasi kelas positif secara tepat. Sementara itu, nilai *recall* yang lebih tinggi menunjukkan bahwa model lebih peka dalam mendeteksi kelas positif. Nilai ini perlu diperhatikan jika mendeteksi sebanyak mungkin kelas positif. Apabila tujuan utama adalah mendapatkan keseimbangan antara *precision* dan *recall*, maka *f1-score* menjadi metrik yang tepat untuk digunakan, semakin besar nilai *f1-score* yang lebih tinggi menunjukkan bahwa model memiliki keseimbangan yang lebih baik antara *precision* dan *recall*.

Hasil rata-rata evaluasi untuk data Halodoc menggunakan kedua metode ditampilkan pada tabel 6. Sementara, hasil rata-rata evaluasi untuk data Alodokter ditunjukkan pada tabel 7. Hasil evaluasi menunjukkan nilai *precision*, *recall*, *f1-score* dan *accuracy*. Rata-rata evaluasi dari kedua metode selanjutnya dibandingkan untuk mengetahui metode dengan kinerja terbaik.

Tabel 6. Hasil Rata-Rata Evaluasi Tiap Metode Pada Dokumen Halodoc

Metode	Precision	Recall	F1-Score	Accuracy
InSet-SVM	**85,92%**	**86,24%**	**85,76%**	**86,24%**
SentiStrength_id-SVM	81,35%	81,21%	81,22%	81,21%

Tabel 7. Hasil Rata-Rata Evaluasi Tiap Metode Pada Dokumen Alodokter

Metode	Precision	Recall	F1-Score	Accuracy
InSet-SVM	**83,86%**	**84,28%**	**83,59%**	**84,28%**
SentiStrength_id-SVM	78,99%	78,94%	78,75%	78,94%

Berdasarkan tabel 6 dan 7, dapat dilihat bahwa metode InSet-SVM memiliki hasil evaluasi yang lebih tinggi dibandingkan dengan metode SentiStrength_id-SVM di tiap kategori *precision*, *recall*, *f1-score* dan *accuracy* baik di dokumen Halodoc maupun Alodokter. Pada dokumen Halodoc, kinerja metode InSet-SVM menghasilkan nilai *precision* 85,92%, nilai *recall* 86,24%, nilai *f1-score* 85,76%, dan nilai *accuracy* sebesar 86,24%. Sementara pada dokumen Alodokter, metode ini menghasilkan nilai *precision* 83,86%, nilai *recall* 84,28%, nilai *f1-score* 83,59%, dan nilai *accuracy* sebesar 84,28%.

4. KESIMPULAN DAN SARAN

Pelabelan data menggunakan kamus sentimen yang berbeda akan menghasilkan label klasifikasi yang berbeda. Pada pelabelan data menggunakan kamus SentiStrength_id, sebagian besar data cenderung diberi label netral, yaitu sebesar 49% untuk data halodoc dan 45% untuk data Alodokter. Sedangkan pada pelabelan menggunakan kamus InSet, sebagian besar data justru diberi label negatif, yaitu masing-masing 59% untuk data Halodoc maupun Alodokter. Hal ini menunjukkan bahwa kamus SentiStrength_id memiliki daftar kata yang kurang beragam dibandingkan dengan kamus InSet. Secara keseluruhan, kamus InSet terbukti memiliki kinerja yang lebih baik. Pada dokumen Halodoc, kinerja metode InSet-SVM menghasilkan nilai *precision* sebesar 85,92%, *recall* sebesar 86,24%, *f1-score* sebesar 85,76%, dan *accuracy* sebesar 86,24%. Sementara pada dokumen Alodokter, metode ini menghasilkan nilai *precision* sebesar 83,86%, *recall* sebesar 84,28%, *f1-score* sebesar 83,59%, dan *accuracy* sebesar 84,28%. Penelitian ini menunjukkan bahwa pemilihan leksikon sebagai sumber pelabel awal memengaruhi distribusi kelas dan kinerja SVM. InSet–SVM memberikan kinerja paling konsisten pada data Halodoc dan Alodokter, dengan *accuracy* yang lebih tinggi daripada SentiStrength_id–SVM.

Penelitian selanjutnya dapat membandingkan SVM dengan Logistic Regression, Random Forest, XGBoost, dan model transformer bahasa Indonesia seperti IndoBERT. Penanganan ketidakseimbangan kelas dapat dikaji melalui class weighting, random oversampling, SMOTE,

atau undersampling yang diterapkan hanya pada data latih di setiap fold. Selain itu, pengembangan leksikon khusus domain telemedicine dan analisis berbasis aspek dapat digunakan untuk menangkap sentimen terhadap aspek-aspek khusus secara lebih komprehensif.

DAFTAR PUSTAKA

[1] C. Kruse and K. Heinemann, "Facilitators and Barriers to the Adoption of Telemedicine During the First Year of COVID-19: Systematic Review.," *J. Med. Internet Res.*, vol. 24, no. 1, p. e31752, 2022, doi: 10.2196/31752.

[2] J. T. L. Anderson, *et al*., "Telehealth adoption during the COVID-19 pandemic: A social media textual and network analysis," *Digit. Heal.*, vol. 8, p. 20552076221090040, Jan. 2022, doi: 10.1177/20552076221090041.

[3] N. Paramita and S. Noviarisanti, "Service Quality Analysis of Mhealth Services Using Text Mining Method : Alodokter and Halodoc," *Int. J. Manag. Financ. Account.*, vol. 2, pp. 1–21, 2021, doi: 10.33093/ijomfa.2021.2.2.1.

[4] T. Eng, M. R. Ibn Nawab, and K. M. Shahiduzzaman, "Improving accuracy of the sentence-level lexicon-based sentiment analysis using machine learning," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, pp. 57–69, 2021, doi: 10.32628/cseit21717.

[5] K. Naithani and Y. P. Raiwani, "Realization of natural language processing and machine learning approaches for text-based sentiment analysis," *Expert Syst.*, vol. 40, no. 5, p. e13114, 2023, doi: https://doi.org/10.1111/exsy.13114.

[6] S. Mutmainah, D. H. Fudholi, and S. Hidayat, "Analisis Sentimen dan Pemodelan Topik Aplikasi Telemedicine Pada Google Play Menggunakan BiLSTM dan LDA," *J. MEDIA Inform. Budidarma*, vol. 7, no. 1, pp. 312–323, 2023, doi: 10.30865/mib.v7i1.5486.

[7] N. R. Wardani and A. Erfina, "Analisis Sentimen Masyarakat Terhadap Layanan Konsultasi Dokter Menggunakan Algoritma Naive Bayes," in *Seminar Nasional Sistem Informasi dan Manajemen Informatika*, 2021, pp. 11–18.

[8] E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, "Implementasi Algoritma Naive Bayes , Support Vector Machine, dan K-Nearest Neighbors Untuk Analisa Sentimen," *Fakt. Exacta*, vol. 14, no. 2, pp. 64–71, 2021, 10.30998/faktorexacta.v14i2.9697.

[9] R. N. Cikania, "Analisis Sentimen Review Pengguna Layanan Telemedicine Halodoc Menggunakan Algoritma Naive Bayes Classifier dan Support Vector Machine,"*Jambura: Jpurnal of Probability and Statistics*, vol. 2, no. 2, pp. 96-104, 2021, doi: 10.34312/jjps.v2i2.11364.

[10] R. L. Pratiwi, Z. I. Alfianti, A. Fauzi, and G. Ginabila, "Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Terhadap Penggunaan True Wireless Stereo (TWS)," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 8, no. 2, pp. 257–268, 2025, doi: 10.36080/skanika.v8i2.3535.

[11] M. Syamsul Hadi, Jihadul Akbar, and Muhammad Fauzi Zulkarnain, "Analisis Sentimen Wisata Air Terjun Di Kabupaten Lombok Tengah Menggunakan Metode Support Vector Machine (SVM)," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 8, no. 2, pp. 318–329, Jul. 2025, doi: 10.36080/skanika.v8i2.3578.

[12] F. S. Yerzi and Y. Sibaroni, "Analisis Sentimen Terhadap Kebijakan Pemerintah Dalam Menangani Covid-19 Dengan Pendekatan Lexicon Based," *eProceedings Eng.*, vol. 8, no. 5, pp. 11354–11366, 2021, [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15616

[13] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017*, 2017, pp. 391–394. doi: 10.1109/IALP.2017.8300625.

[14] D. H. Wahid and A. SN, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol.

10, no. 2, p. 207, 2016, doi: 10.22146/ijccs.16625.
[15] M. A. Palomino and F. Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Applied Sciences*, vol. 12, no. 17. p. 8765, 2022. doi: 10.3390/app12178765.
[16] O. I. Gifari, M. Adha, I. R. Hendrawan, and F. F. S. Durrand, "Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022, doi: 10.46229/jifotech.v2i1.330.
[17] M. Setiawan, I. Gunadi, and G. Indrawan, "Klasifikasi Pelayanan Kesehatan Berdasarkan Data Sentimen Pelayanan Kesehatan menggunakan Multiclass Support Vector Machine," *J. Sist. dan Inform.*, vol. 17, pp. 47–54, 2023, doi: 10.30864/jsi.v17i1.512.
[18] N. Ririanti and A. Purwinarko, "Implementation of Support Vector Machine Algorithm with Correlation-Based Feature Selection and Term Frequency Inverse Document Frequency for Sentiment Analysis Review Hotel," *Sci. J. Informatics*, vol. 8, pp. 297–303, 2021, doi: 10.15294/sji.v8i2.29992.