

Analisis Sentimen Multi-Aspek pada Ulasan Aplikasi MySiloam Menggunakan Pipeline *BERTopic* dengan Perbandingan Algoritma *Clustering*

Jihan Hasna Iftinan, Eka Dyar Wahyuni, Reisa Permatasari

Program Studi Sistem Informasi, Universitas Pembangunan Nasional “Veteran” Jawa Timur Jl. Raya Rungkut Madya, Gunung Anyar Surabaya Jawa Timur
Email: jihaniftinan.0110@gmail.com

ABSTRAK

Tingginya volume ulasan pengguna aplikasi kesehatan digital belum dimanfaatkan secara optimal untuk memahami aspek spesifik yang memengaruhi pengalaman pengguna. Penelitian ini bertujuan menganalisis sentimen berbasis multi-aspek pada ulasan aplikasi MySiloam menggunakan metode *BERTopic* untuk ekstraksi aspek dan SVM *One-vs-One* untuk klasifikasi sentimen. Sebanyak 2.657 ulasan dikumpulkan dari Google Play Store dan App Store rentang 2019–2025, disaring menjadi 1.699 ulasan setelah *preprocessing*. *BERTopic* dijalankan dengan perbandingan tiga algoritma *clustering* (HDBSCAN, BIRCH, K-Means) dan klasifikasi sentimen dibandingkan dalam dua skenario yaitu pendekatan dua tahap dan klasifikasi gabungan. K-Means dengan *stemming* menghasilkan tiga aspek layanan utama dengan kualitas topik terbaik, sementara pendekatan dua tahap menghasilkan *F1-score* tertinggi 89,53%, membuktikan bahwa kombinasi *BERTopic* dan SVM OvO efektif sebagai solusi otomatis analisis sentimen berbasis aspek pada ulasan aplikasi kesehatan digital berbahasa Indonesia.

Kata Kunci: *Aspect-Based Sentiment Analysis, BERTopic, Support Vector Machine, MySiloam, Topic Modeling*

ABSTRACT

The high volume of digital health application user reviews has not been optimally utilized to understand specific aspects affecting user experience. This study aims to analyze multi-aspect sentiment on MySiloam application reviews using BERTopic for aspect extraction and SVM One-vs-One for sentiment classification. A total of 2,657 reviews were collected from Google Play Store and App Store spanning 2019–2025, filtered to 1,699 reviews after preprocessing. BERTopic was applied with a comparison of three clustering algorithms (HDBSCAN, BIRCH, K-Means) and sentiment classification was compared across two scenarios: a two-stage approach and a joint classification approach. K-Means with stemming produced three main service aspects with the best topic quality, while the two-stage approach achieved the highest F1-score of 89.53%, demonstrating that the combination of BERTopic and SVM OvO is effective as an automated solution for aspect-based sentiment analysis on Indonesian digital health application reviews.

Keywords: *Aspect-Based Sentiment Analysis, BERTopic, Support Vector Machine, MySiloam, Topic Modeling*

1. PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi telah mengubah lanskap layanan kesehatan secara mendasar, mendorong transformasi dari sistem konvensional menuju ekosistem digital yang lebih terintegrasi. Di Indonesia, digitalisasi sektor kesehatan berkembang dengan pesat: sekitar 57% masyarakat Indonesia telah memanfaatkan aplikasi kesehatan digital, menjadikan Indonesia sebagai pasar ketiga terbesar di dunia dalam adopsi teknologi kesehatan (Databokss, 2024). Relevansi digitalisasi ini diperkuat oleh temuan Badan Pusat Statistik (BPS) tahun 2025 yang mengungkapkan bahwa setiap kenaikan 1% dalam tingkat digitalisasi berpotensi meningkatkan indeks kualitas kesehatan sebesar 0,164 poin, menegaskan peran strategis transformasi digital dalam pembangunan sosial-ekonomi nasional (Badan Pusat Statistik, 2025). Pergeseran perilaku pengguna pun semakin nyata, terutama di kalangan generasi muda; survei Indonesia Sentinel melaporkan bahwa 69% Generasi Z dan Milenial di Indonesia secara aktif memanfaatkan layanan telehealth untuk memenuhi kebutuhan kesehatan mereka sehari-hari (Indonesia Sentinel, 2025).

Di sektor swasta, Siloam Hospitals Group sebagai jaringan rumah sakit terbesar di Indonesia dengan 41 unit dan lebih dari 7.900 tempat tidur turut mengambil peran dalam transformasi ini melalui peluncuran aplikasi MySiloam (PT Siloam International Hospitals Tbk, 2025). Hingga tahun 2025, MySiloam telah diunduh lebih dari satu juta kali dan memperoleh lebih dari 33 ribu ulasan pengguna di Google Play Store, mencerminkan tingginya tingkat adopsi sekaligus menyimpan potensi besar untuk penggalan wawasan berbasis data.

Meskipun volume ulasan yang sangat besar tersebut merupakan sumber informasi berharga, analisis manual terhadapnya sangat tidak efisien. Penelitian sebelumnya oleh (Lia et al., 2025) telah menganalisis 1.995 ulasan MySiloam menggunakan metode Naïve Bayes dengan representasi TF-IDF dan menghasilkan akurasi sekitar 86%. Dari analisis tersebut, ditemukan bahwa mayoritas pengguna memberikan ulasan positif mengenai fitur dan kemudahan penggunaan aplikasi, meskipun terdapat beberapa kritik terkait performa aplikasi. Namun, pendekatan tersebut masih terbatas pada klasifikasi sentimen umum (positif dan negatif) tanpa mampu mengidentifikasi aspek-aspek spesifik yang membentuk persepsi pengguna.

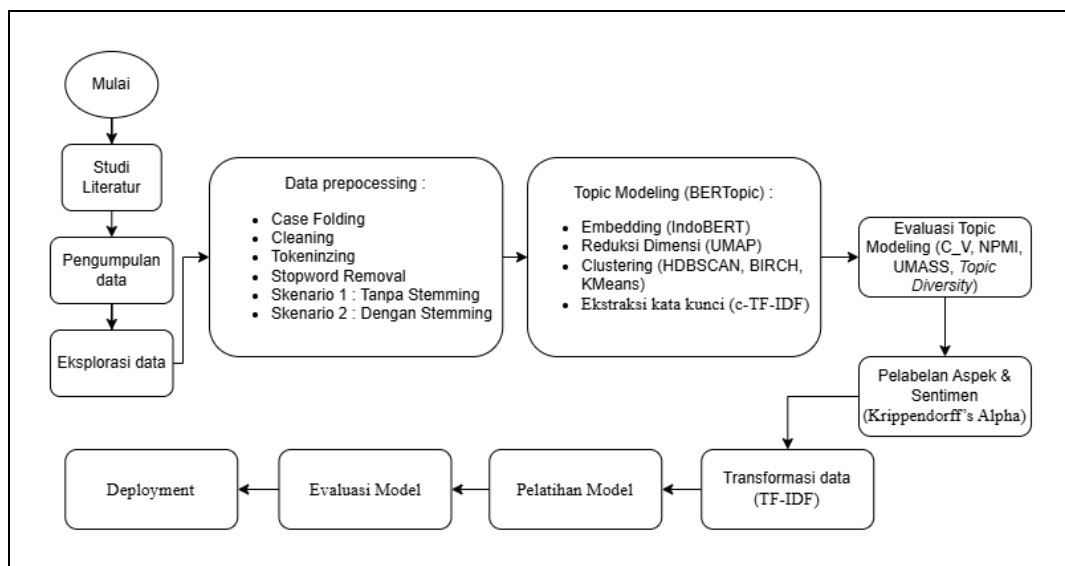
Studi serupa pada aplikasi kesehatan lain di Indonesia juga menghadapi keterbatasan yang sama. (Girsang, 2024) menganalisis ulasan pengguna aplikasi Alodokter dan Halodoc menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel linier, penelitian tersebut berhasil mengidentifikasi pola sentimen yang dominan, namun tetap terbatas pada level klasifikasi sentimen menyeluruh tanpa mengurai aspek layanan yang spesifik. Sementara itu, penelitian pada tiga aplikasi layanan kesehatan populer seperti Halodoc, Alodokter, dan Klikdokter menggunakan pendekatan IndoBERT dengan fine-tuning mencapai akurasi 96%, mengungguli metode sebelumnya (Imaduddin et al., 2023). Meski unggul dari sisi akurasi, seluruh penelitian tersebut masih beroperasi pada level sentimen dokumen secara keseluruhan. Padahal, ulasan pengguna aplikasi kesehatan termasuk MySiloam menunjukkan variasi yang signifikan, mulai dari pujian terhadap kemudahan penggunaan hingga kritik terhadap keterbatasan integrasi antarplatform dan ketidakkonsistenan data rekam medis antarcabang. Kesenjangan metodologis ini menegaskan perlunya pendekatan yang lebih mendalam.

Untuk menjawab keterbatasan tersebut, penelitian ini mengadopsi pendekatan *Aspect-Based Sentiment Analysis* (ABSA) sebagai metode analisis sentimen yang bersifat fine-grained, yakni mampu mengidentifikasi aspek spesifik beserta polaritas opini yang menyertainya secara bersamaan (Hua et al., 2024). Sejalan dengan itu, penelitian (Sari Pranasti, 2025) pada aplikasi layanan medis digital Indonesia menunjukkan bahwa penerapan ABSA dengan SVM berhasil mengungkap aspek-aspek seperti performa, fitur, dan harga sebagai faktor penentu persepsi pengguna, membuktikan relevansi pendekatan ini dalam domain serupa.

Ekstraksi aspek secara otomatis dari data ulasan berskala besar dalam penelitian ini dilakukan menggunakan *BERTopic*, sebuah metode *topic modeling* berbasis transformer embedding yang menggabungkan UMAP untuk reduksi dimensi dan *class-based* TF-IDF (c-TF-IDF) untuk representasi topik (Grootendorst, 2022). Dibandingkan dengan pendekatan klasik seperti LDA, NMF, dan LSA, *BERTopic* terbukti menghasilkan topik dengan koherensi semantik lebih tinggi khususnya dalam domain medis (Egger & Yu, 2022), serta lebih adaptif terhadap variasi konteks kalimat (Ma et al., 2025).

Selanjutnya, klasifikasi sentimen pada setiap aspek yang teridentifikasi dilakukan menggunakan algoritma *Support Vector Machine* (SVM), yang dipilih karena kemampuannya dalam menangani data berdimensi tinggi dengan margin klasifikasi optimal melalui konstruksi hyperplane pemisah, sehingga menghasilkan performa yang unggul pada data teks yang kompleks dan tidak seimbang. SVM memiliki keunggulan dalam memisahkan data sentimen positif dan negatif serta menghasilkan akurasi tinggi dalam klasifikasi sentimen, terutama pada data yang bersifat *linearly separable*, dengan cara memetakan data ke dimensi lebih tinggi untuk menemukan hyperplane pemisah yang ideal (Badri et al., 2025). Dengan demikian, penelitian ini berkontribusi dalam mengisi kesenjangan metodologis pada analisis ulasan aplikasi kesehatan di Indonesia melalui kombinasi *BERTopic* dan SVM dalam kerangka ABSA, dengan objek kajian aplikasi MySiloam yang selama ini belum dieksplorasi secara mendalam pada level aspek.

2. METODE



Gambar 1: Alur Penelitian

Penelitian ini dilaksanakan melalui tahapan sistematis sebagaimana ditunjukkan pada Gambar 1, meliputi studi literatur, pengumpulan data, eksplorasi data, data *preprocessing*, *topic modeling*, evaluasi *topic modeling*, pelabelan aspek dan sentimen, pembagian data, transformasi data, pelatihan model, evaluasi model, dan *deployment*.

Studi Literatur

Studi literatur dilakukan melalui penelusuran jurnal ilmiah nasional dan internasional yang relevan dengan topik *Aspect-Based Sentiment Analysis* (ABSA), *BERTopic*, algoritma klasterisasi (HDBSCAN, BIRCH, K-Means), serta SVM untuk klasifikasi sentimen. Hasil kajian ini menjadi landasan teoritis dan metodologis dalam perancangan penelitian.

Pengumpulan Data

Data berupa ulasan pengguna aplikasi MySiloam dikumpulkan dari Google Play Store dan App Store menggunakan library `google_play_scraper` pada Python di platform Google Colab, Sedangkan untuk data dari App Store, digunakan kombinasi library `requests`, `pandas`, `datetime`, dan `time` untuk melakukan web *scraping* terhadap API publik milik Apple, mencakup rentang waktu 2019–2025. Seluruh hasil *scraping* disimpan dalam format CSV sebagai bahan dasar analisis.

Eksplorasi Data

Eksplorasi data dilakukan untuk memahami karakteristik awal dataset, meliputi analisis distribusi rating bintang, panjang teks ulasan, dan frekuensi kemunculan kata. Visualisasi WordCloud digunakan untuk menggambarkan tema dominan dalam dataset menggunakan pustaka `wordcloud` pada Python (Ren et al., 2024).

Data Preprocessing

Tahap *preprocessing* bertujuan membersihkan dan menstandarisasi data teks mentah melalui empat tahap berurutan: *case folding* (konversi teks ke huruf kecil), *cleaning* (penghapusan angka, tanda baca, emoji, dan URL menggunakan regex), *tokenizing* (pemecahan teks menjadi unit kata), dan *stopword removal* (penghapusan kata umum menggunakan kamus Sastrawi). Setelah keempat tahap tersebut, data dibagi menjadi dua skenario: Skenario 1 tanpa *stemming* dan Skenario 2 dengan *stemming* menggunakan algoritma Sastrawi, guna membandingkan pengaruh reduksi morfologi terhadap kualitas *topic modeling* (Aufar et al., 2023).

Topic Modeling

Ekstraksi aspek dilakukan menggunakan *BERTopic* melalui empat tahap. Pertama, embedding teks menggunakan IndoBERT yang menghasilkan vektor

berdimensi 768. Kedua, reduksi dimensi menggunakan UMAP untuk memproyeksikan vektor ke ruang berdimensi lebih rendah sambil mempertahankan struktur semantik. Ketiga, *clustering* menggunakan tiga algoritma yang dibandingkan: HDBSCAN, BIRCH, dan K-Means. Keempat, ekstraksi kata kunci menggunakan *class-based* TF-IDF (c-TF-IDF) yang menghitung bobot kata berdasarkan distribusinya per kluster, berbeda dengan TF-IDF konvensional yang bekerja per dokumen (Muhajir et al., 2024).

Evaluasi *Topic Modeling*

Kualitas topik yang dihasilkan dievaluasi menggunakan empat metrik, yaitu *Coherence* (C_v), UMass, *Normalized Pointwise Mutual Information* (NPMI) dan *Topic Diversity*. Kombinasi dua skenario *preprocessing* dan tiga algoritma *clustering* menghasilkan enam skenario evaluasi. Model dengan nilai C_v tertinggi dipilih sebagai model aspek terbaik untuk digunakan pada tahap klasifikasi sentimen (Muhajir et al., 2024).

Pelabelan Aspek dan Sentimen

Pelabelan aspek dan sentimen dilakukan secara manual oleh tiga annotator independen. Setiap ulasan dapat dipecah menjadi beberapa segmen apabila membahas lebih dari satu aspek layanan (*multi-aspect*). Untuk memastikan konsistensi antar pelabel, dilakukan evaluasi menggunakan metrik Krippendorff's Alpha, dengan nilai $\alpha \geq 0,67$ yang menunjukkan tingkat kesepakatan yang dapat diterima (Krippendorff, 2011).

Pembagian Data

Dataset hasil pelabelan dibagi menjadi data latih (90%) dan data uji (10%) menggunakan teknik stratified sampling agar distribusi kelas sentimen positif, negatif, dan netral tetap proporsional di kedua subset (Wibowo et al., 2024).

Transformasi Data

Transformasi data dilakukan menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF) sebagai teknik representasi utama. TF-IDF dipilih karena mampu memberikan bobot lebih besar pada kata-kata yang memiliki nilai pembeda tinggi antar dokumen, sehingga membantu model dalam mengenali kata-kata yang berkontribusi penting terhadap penentuan sentimen (Muhajir et al., 2024).

Pelatihan Model

Klasifikasi sentimen pada setiap aspek yang teridentifikasi dilakukan menggunakan algoritma *Support Vector Machine* (SVM) OvO. SVM dipilih karena kemampuannya menangani data berdimensi tinggi dengan margin klasifikasi optimal melalui konstruksi hyperplane pemisah, sehingga menghasilkan performa unggul pada data teks yang kompleks. Parameter model dioptimalkan melalui penyesuaian kernel dan *regularization parameter* (C) untuk mencegah *overfitting* (Setiono, 2022) (Du et al., 2024).

Evaluasi Model

Performa model dievaluasi menggunakan empat metrik utama berdasarkan confusion matrix, yaitu akurasi, presisi, *recall*, dan *F1-score*. Metrik-metrik ini dipilih karena mampu memberikan gambaran menyeluruh terhadap performa model dari berbagai perspektif, terutama pada kondisi distribusi kelas yang tidak seimbang (Shin et al., 2023).

Deployment Model

Model yang telah divalidasi diimplementasikan ke dalam aplikasi berbasis web menggunakan framework Flask sebagai backend dan HTML/CSS sebagai frontend. Integrasi Chart.js digunakan untuk menampilkan hasil analisis sentimen per aspek dalam bentuk grafik distribusi yang interaktif, sehingga memudahkan interpretasi oleh tim pengembang MySiloam maupun pihak manajemen rumah sakit.

3. HASIL DAN PEMBAHASAN

Data penelitian ini diperoleh dari dua platform utama, yaitu Google Play Store dan App Store. Proses *scraping* menghasilkan sebanyak 2.376 ulasan dari Google Play Store dan 281 ulasan dari App Store, sehingga total keseluruhan data yang terkumpul adalah 2.657 ulasan pengguna aplikasi MySiloam dalam rentang waktu 2019–2025. Seluruh data disimpan dalam format CSV sebagai bahan dasar untuk tahap analisis selanjutnya. Berdasarkan hasil distribusi jumlah kata, ditemukan bahwa dari total 2.657 ulasan, sebanyak 958 ulasan (36,06%) memiliki kurang dari 3 kata dan dianggap tidak memberikan informasi yang cukup untuk dianalisis sebagai penelitian multi aspek. Oleh karena itu, seluruh ulasan dengan jumlah kata kurang dari 3 dihapus pada tahap

cleaning, sehingga dataset yang tersisa dan digunakan untuk analisis selanjutnya berjumlah 1.699 ulasan. Proses *preprocessing* secara keseluruhan terdiri dari empat tahap berurutan, yaitu *case folding*, *cleaning*, *tokenizing*, dan *stopword removal*. Untuk menggambarkan proses tersebut secara konkret, berikut Tabel 1 sebagai contoh transformasi teks pada setiap tahapan menggunakan kalimat "Aplikasi lengkap fitur andalan dan Sangat Informatif":

Tabel 1. Hasil Data *Preprocessing*

Tahap	Sebelum	Sesudah
<i>Case Folding</i>	Aplikasi lengkap fitur andalan dan Sangat Informatif 👍	aplikasi lengkap fitur andalan dan sangat informatif 👍
<i>Cleaning</i>	aplikasi lengkap fitur andalan dan sangat informatif 👍	aplikasi lengkap fitur andalan dan sangat informatif
<i>Tokenizing</i>	aplikasi lengkap fitur andalan dan sangat informatif	['aplikasi', 'lengkap', 'fitur', 'andalan', 'dan', 'sangat', 'informatif']
<i>Stopword Removal</i>	['aplikasi', 'lengkap', 'fitur', 'andalan', ' dan ', 'sangat', 'informatif']	['aplikasi', 'lengkap', 'fitur', 'andalan', 'informatif']

Sumber: Data Sekunder, diolah (2026)

Setelah keempat tahap tersebut, data dibagi menjadi dua skenario untuk membandingkan pengaruh *stemming* terhadap kualitas *topic modeling*:

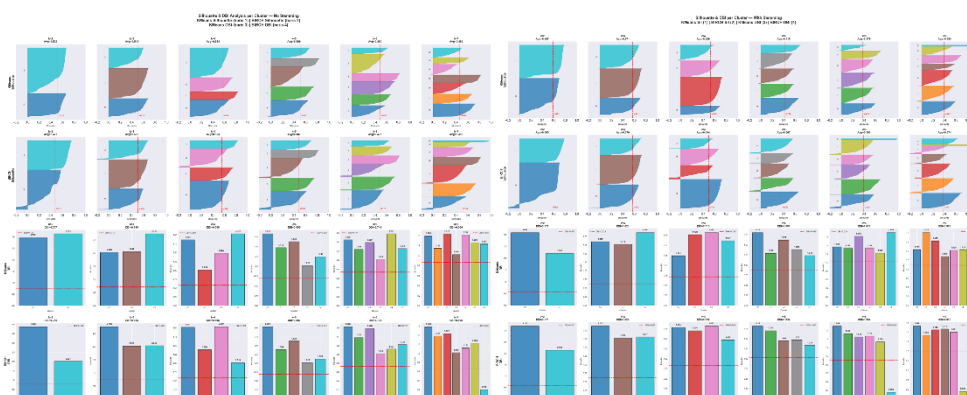
Tabel 2. Hasil *Stemming*

Skenario	Hasil Akhir
Tanpa <i>Stemming</i>	['aplikasi', 'lengkap', 'fitur', ' andalan ', 'informatif']
Dengan <i>Stemming</i>	['aplikasi', 'lengkap', 'fitur', ' andal ', 'informasi']

Sumber: Data Sekunder, diolah (2026)

Pada Tabel 2 perbedaan utama antara kedua skenario terlihat pada kata "andalan" yang menjadi "andal" dan "informatif" yang menjadi "informasi" setelah proses *stemming*. Hal ini menunjukkan bahwa *stemming* menyederhanakan variasi morfologi kata. Pada tahap *clustering BERTopic*, tiga algoritma dibandingkan yaitu HDBSCAN, BIRCH, dan K-Means. Khusus untuk algoritma BIRCH dan K-Means yang

memerlukan penentuan jumlah kluster (k) secara manual, dilakukan analisis *Silhouette Score* dan *Davies-Bouldin Index (DBI)* untuk menentukan nilai k optimal.



Gambar 2: *Silhouette Score* dan *Davies-Bouldin Index (DBI)* skenario *stemming* & tanpa *stemming*

Berdasarkan visualisasi pada Gambar 2 (tanpa *stemming* dan dengan *stemming*) nilai $k=3$ dipilih sebagai k optimal untuk kedua algoritma pada kedua skenario. Keputusan ini didasarkan pada dua pertimbangan yaitu nilai *Silhouette Score* pada $k=3$ menunjukkan keseimbangan yang baik antara kekompakan dan pemisahan antar kluster dan nilai *DBI* pada $k=3$ relatif lebih rendah dibandingkan k lainnya dengan selisih nilai negatif antarkluster yang paling minimal, mengindikasikan kluster yang lebih homogen dan terpisah dengan baik. Hasil *BERTopic* pada kedua skenario *preprocessing* seperti pada Tabel 3 berikut.

Tabel 3. Hasil *Topic Modeling*

Skenario	Cluster	Jumlah Topic
Tanpa <i>Stemming</i>	HDBSCAN	5 topik dengan <i>noise</i> sebesar 42
	BIRCH	3 topik
	K-MEANS	3 topik
Dengan <i>Stemming</i>	HDBSCAN	9 topik dengan <i>noise</i> lebih besar yaitu 88 dokumen
	BIRCH	3 topik
	K-MEANS	3 topik

Sumber: Data Sekunder, diolah (2026)

Perbedaan jumlah topik HDBSCAN antara dua skenario menunjukkan bahwa *stemming* memengaruhi distribusi semantik dokumen sehingga HDBSCAN mendeteksi lebih banyak kelompok topik yang terpisah. Selanjutnya evaluasi kualitas topik dilakukan menggunakan 4 metrik, yaitu *Coherence* (C_v), *UMass*, *NPMI*, *Topic Diversity*.

Tabel 4. Hasil Perbandingan Evaluasi *BERTopic* tanpa *stemming* dan dengan *stemming*

No	Algoritma	Skenario	N Topik	C_v	UMass	NPMI	<i>Topic Diversity</i>
1	HDBSCAN	Tanpa <i>Stemming</i>	5	0,2348	-16,1669	-0,3807	0,9800
2	HDBSCAN	Dengan <i>Stemming</i>	9	0,4108	-13,2922	NaN	0,9889
3	BIRCH	Tanpa <i>Stemming</i>	3	0,3167	-13,4242	-0,3477	1,0000
4	BIRCH	Dengan <i>Stemming</i>	3	0,2858	-14,5805	-0,4137	1,0000
5	K-Means	Tanpa <i>Stemming</i>	3	0,3917	-4,3617	-0,0423	0,9000
6	K-Means	Dengan <i>Stemming</i>	3	0,4113	-3,8012	-0,0200	0,9000

Sumber: Data Sekunder, diolah (2026)

Berdasarkan hasil evaluasi pada Tabel 4, model K-Means dengan *stemming* dipilih sebagai model terbaik dengan pertimbangan sebagai berikut: Pertama, nilai C_v tertinggi sebesar 0,4113 mendekati threshold 0,5 yang mengindikasikan koherensi topik yang cukup baik dibandingkan seluruh model lainnya. Kedua, nilai UMass terbaik sebesar -4,3434, paling mendekati 0 di antara semua model, menunjukkan bahwa kata-kata dalam satu topik sering muncul bersama secara konsisten dalam korpus. Ketiga, nilai NPMI sebesar 0,0200 merupakan yang tertinggi dan satu-satunya yang tidak negatif, nilai HDBSCAN pada skenario *stemming* tidak muncul karena banyaknya noise yang dihasilkan, sehingga K-MEANS dengan *stemming* mengindikasikan hubungan semantik antarkata dalam topik yang paling kuat. Keempat, nilai *Topic Diversity* sebesar 0,9000 di atas threshold 0,6 menunjukkan bahwa topik-topik yang dihasilkan cukup beragam dan tidak tumpang tindih. Secara keseluruhan, hasil ini juga mengkonfirmasi bahwa penggunaan *stemming* pada data ulasan MySiloam memberikan dampak positif terhadap kualitas *topic modeling*, khususnya pada algoritma K-Means,

karena reduksi morfologi membantu menyatukan variasi kata yang bermakna sama sehingga menghasilkan kluster yang lebih koheren. Berdasarkan hasil *topic modeling* K-Means dengan *stemming*, diperoleh tiga topik yang kemudian diinterpretasikan sebagai aspek layanan aplikasi MySiloam sebagaimana seperti pada Tabel 5 berikut.

Tabel 5. Interpretasi Hasil *Topic Modeling*

Topik	Label Aspek	Kata Kunci Dominan
1	Fitur Medis Aplikasi	dokter, sakit, rumah, pasien, antri, janji, obat, jadwal
2	Pelayanan & Kepuasan	layan, baik, bagus, bantu, cepat, mudah, ramah, puas
3	Teknis Aplikasi	aplikasi, daftar, login, error, booking, update, sulit

Sumber: Data Sekunder, diolah (2026)

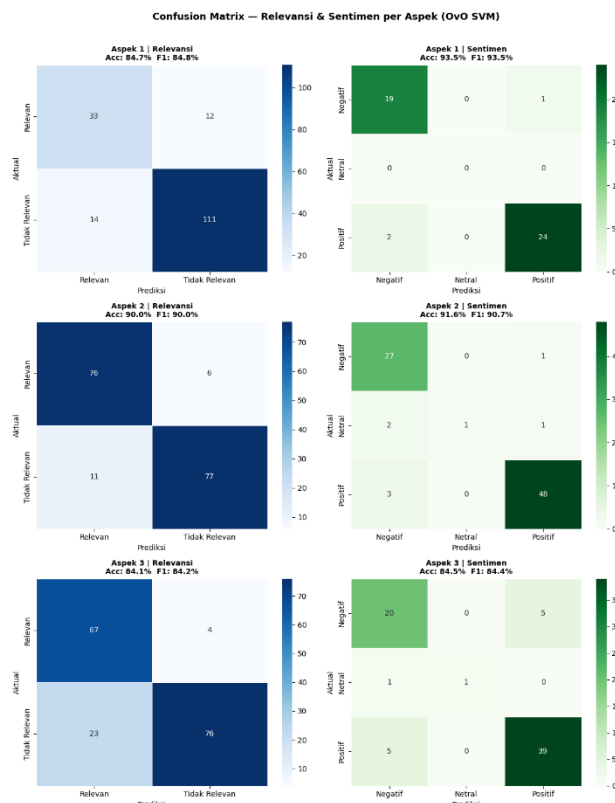
Setelah aspek teridentifikasi, dilakukan pelabelan sentimen secara manual oleh tiga annotator independen. Konsistensi antar pelabel diukur menggunakan metrik Krippendorff's Alpha dengan hasil sebesar 0,8816 menunjukkan tingkat kesepakatan yang sangat baik (*reliable*) antar ketiga annotator, jauh di atas ambang batas minimum $\alpha \geq 0,67$ yang ditetapkan sebagai syarat keandalan pelabelan. Hal ini mengindikasikan bahwa dataset berlabel yang dihasilkan memiliki kualitas yang konsisten dan dapat dipercaya sebagai ground truth untuk tahap pelatihan model. Kemudian klasifikasi sentimen dilakukan menggunakan algoritma SVM dengan pendekatan *One-vs-One* (OvO) untuk setiap aspek yang telah teridentifikasi. Model dilatih dengan dua tahap klasifikasi per aspek, yaitu klasifikasi relevansi (Relevan / Tidak Relevan) dan klasifikasi sentimen (Negatif / Netral / Positif). Hasil evaluasi model ditunjukkan Tabel 6.

Tabel 6. Hasil Evaluasi Model

Aspek	Model	Accuracy	Precision	Recall	F1-Score
Fitur Medis Aplikasi	Relevansi	84,71%	84,94%	84,71%	84,81%
	Sentimen	93,48%	93,60%	93,48%	93,49%
Pelayanan & Kepuasan	Relevansi	90,00%	90,16%	90,00%	90,00%
	Sentimen	91,57%	92,27%	91,57%	90,69%
Teknis Aplikasi	Relevansi	84,12%	86,42%	84,12%	84,21%
	Sentimen	84,51%	84,83%	84,51%	84,42%

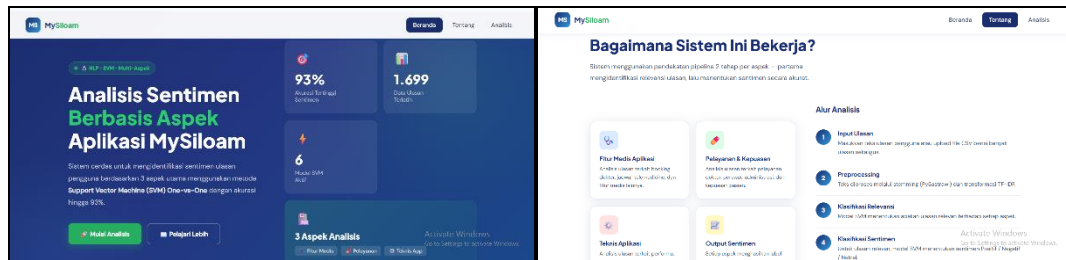
Sumber: Data Sekunder, diolah (2026)

Berdasarkan hasil evaluasi, model menunjukkan performa yang baik pada seluruh aspek. Aspek Fitur Medis Aplikasi mencapai akurasi sentimen tertinggi sebesar 93,48% dengan *F1-score* 93,49%, menunjukkan bahwa model sangat mampu membedakan ulasan positif dan negatif terkait fitur medis seperti janji temu dan antrian. Aspek Pelayanan & Kepuasan memperoleh akurasi relevansi tertinggi sebesar 90,00%, mengindikasikan bahwa model paling akurat dalam mendeteksi ulasan yang membahas topik pelayanan. Sementara aspek Teknis Aplikasi memiliki performa paling rendah dengan akurasi sentimen 84,51%, yang dapat dipahami mengingat ulasan teknis sering mengandung bahasa yang lebih ambigu dan bervariasi.



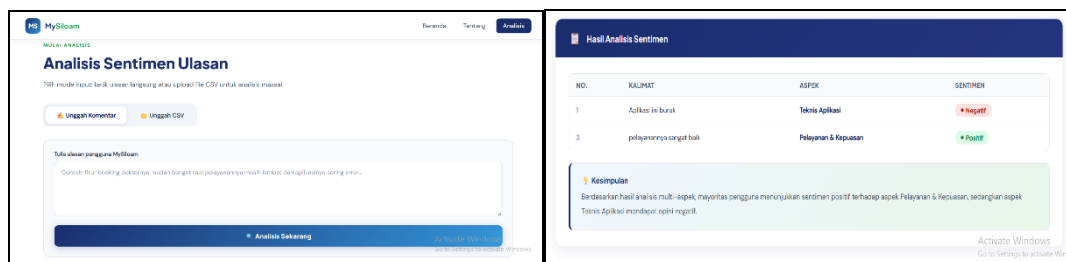
Gambar 3: *Confusion Matrix*

Hasil *confusion matrix* pada Gambar 3 memperlihatkan bahwa kesalahan prediksi terbesar terjadi pada kelas Netral, yang konsisten di semua aspek. Hal ini disebabkan oleh jumlah sampel kelas Netral yang jauh lebih sedikit dibandingkan kelas Positif dan Negatif, sehingga model kurang memiliki referensi yang cukup untuk mengenali pola sentimen netral secara akurat. Secara keseluruhan, pendekatan dua tahap klasifikasi (relevansi + sentimen) dengan SVM OvO terbukti efektif dalam menghasilkan prediksi yang granular dan akurat. Model final yang digunakan pada tahap *deployment* adalah K-Means dengan *stemming* untuk komponen *topic modeling* dan Skenario 1 (pendekatan dua tahap) untuk komponen klasifikasi sentimen. Sistem yang dibangun mampu menerima input berupa teks ulasan pengguna aplikasi MySiloam dan menghasilkan output berupa prediksi aspek layanan beserta polaritas sentimennya secara otomatis. Sistem ini diharapkan dapat memberikan wawasan yang berguna bagi pihak pengembang MySiloam dalam memahami persepsi pengguna terhadap tiga aspek layanan utama, yaitu Fitur Medis Aplikasi, Pelayanan & Kepuasan, dan Teknis Aplikasi, sebagai dasar pengambilan keputusan untuk peningkatan layanan aplikasi secara berkelanjutan. Berikut tampilan dari website seperti pada Gambar 4, 5, 6, dan 7.



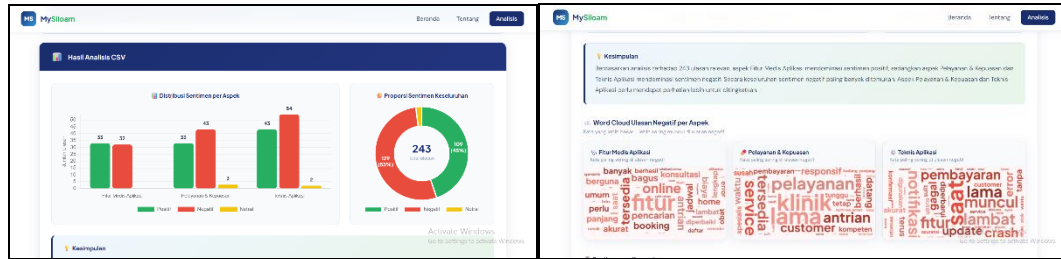
Gambar 4: Beranda dan Informasi Website

Gambar 4 menampilkan halaman beranda sistem analisis sentimen berbasis aspek aplikasi MySiloam. Halaman ini menyajikan ringkasan informasi penelitian, statistik utama sistem seperti jumlah data ulasan dan jumlah aspek yang diidentifikasi, serta tombol navigasi menuju fitur analisis. Selain itu, terdapat halaman informasi yang menjelaskan alur kerja sistem secara visual mulai dari *input* ulasan, *preprocessing*, hingga klasifikasi sentimen per aspek, sehingga pengguna dapat memahami cara kerja sistem sebelum menggunakannya.



Gambar 5: Unggah komentar dan csv

Gambar 5 menampilkan halaman analisis sistem yang menyediakan dua metode input data yaitu pengetikan ulasan secara langsung (*input* teks manual) dan unggah file CSV untuk analisis data dalam jumlah banyak sekaligus. Setelah analisis dijalankan, sistem menampilkan hasil klasifikasi dalam bentuk tabel yang memuat kolom kalimat, aspek, dan sentimen untuk setiap segmen ulasan, disertai kesimpulan otomatis yang merangkum distribusi sentimen secara keseluruhan.



Gambar 6: Visualisasi

Gambar 6 menampilkan halaman visualisasi hasil analisis CSV yang menyajikan tiga komponen utama. Pertama, grafik batang distribusi sentimen per aspek yang menunjukkan perbandingan jumlah ulasan positif, negatif, dan netral pada setiap aspek layanan. Kedua, diagram lingkaran proporsi sentimen keseluruhan. Ketiga, *word cloud* per aspek yang menampilkan kata kunci dominan pada masing-masing aspek layanan, Fitur Medis Aplikasi, Pelayanan & Kepuasan, dan Teknis Aplikasi, sehingga pengguna dapat dengan mudah mengidentifikasi topik yang paling banyak dibahas pengguna pada setiap aspek.

NO.	ULASAN ASLI	KALIMAT	ASPEK	SENTIMEN
1	fitur booking mudah tapi pelayanan dokter sangat mengcewakan	fitur booking mudah pelayanan dokter sangat mengcewakan	Fitur Medis Aplikasi Pelayanan & Kepuasan	Positif Positif
2	aplikasi cepat dan responsif namun pembayaran sering gagal	aplikasi cepat responsif pembayaran sering gagal	Teknis Aplikasi Teknis Aplikasi Teknis Aplikasi	Positif Positif Negatif
3	jadwal dokter lengkap tapi aplikasi sering crash saat dibuka	jadwal dokter lengkap aplikasi sering crash saat dibuka	Fitur Medis Aplikasi Teknis Aplikasi	Negatif Negatif
4	pelayanan ramah namun fitur pencarian dokter masih lambat	pelayanan ramah fitur pencarian dokter masih lambat	Pelayanan & Kepuasan Fitur Medis Aplikasi	Positif Negatif

Gambar 7: Tabel Ulasan Sentimen

Pada Gambar 7 kesalahan klasifikasi pada contoh di atas merupakan keterbatasan yang wajar dalam sistem machine learning, di mana performa model sepenuhnya terikat pada pola yang dipelajari dari data pelatihan. Model SVM yang digunakan mempelajari asosiasi antara fitur teks dan label sentimen berdasarkan distribusi data latih, sehingga ketika dihadapkan pada kalimat kontradiktif yang polanya jarang muncul dalam data pelatihan, model cenderung salah dalam menginterpretasikan konteks kalimat secara keseluruhan. Dengan demikian, tidak ada model machine learning yang sempurna, semakin beragam dan representatif data pelatihan yang digunakan, semakin baik kemampuan model dalam menangani variasi pola bahasa pada data baru.

4. SIMPULAN

Penelitian ini berhasil mengimplementasikan pendekatan *Aspect-Based Sentiment Analysis* (ABSA) pada ulasan pengguna aplikasi MySiloam menggunakan kombinasi *BERTopic* untuk ekstraksi aspek dan SVM *One-vs-One* (OvO) untuk klasifikasi sentimen. Berdasarkan hasil penelitian, diperoleh tiga kesimpulan utama sesuai tujuan penelitian. Pertama, dari perbandingan tiga algoritma *clustering* (HDBSCAN, BIRCH, dan K-Means) dalam *pipeline BERTopic*, algoritma K-Means dengan skenario *stemming* terbukti menghasilkan topik dengan kualitas terbaik, ditunjukkan oleh nilai C_v tertinggi (0,4113), UMass terbaik (-4,3434), NPMI (-0,0206), dan *Topic Diversity* (0,9000). Kedua, model SVM OvO yang dibangun menunjukkan performa klasifikasi yang baik pada seluruh aspek. Akurasi sentimen tertinggi dicapai pada aspek Fitur Medis Aplikasi sebesar 93,48% (F1-score 93,49%), diikuti Pelayanan & Kepuasan sebesar 91,57% (F1-score 90,69%), dan Teknis Aplikasi sebesar 84,51% (F1-score 84,42%). Pendekatan dua tahap klasifikasi (relevansi dan sentimen) terbukti lebih efektif dibandingkan klasifikasi sentimen tunggal karena mampu menyaring ulasan yang tidak relevan terlebih dahulu sehingga menghasilkan prediksi yang lebih granular dan akurat.

Penelitian ini memiliki keterbatasan pada rendahnya performa klasifikasi kelas Netral akibat ketidakseimbangan distribusi data. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi penggunaan model transformer seperti IndoBERT *fine-tuning* sebagai pengganti SVM guna mengatasi keterbatasan tersebut, serta memperluas cakupan data dengan mempertimbangkan ulasan multibahasa dari App Store untuk meningkatkan representativitas analisis.

5. DAFTAR PUSTAKA

- Aufar, A. F., Mochamad Alfian Rosid, Eviyanti, A., & Astutik, I. R. I. (2023). Optimizing Text *Preprocessing* for Accurate Sentiment Analysis on E-Wallet Reviews. *JICTE (Journal of Information and Computer Technology Education)*, 7(2), 42–50. <https://doi.org/10.21070/jicte.v7i2.1650>
- Badan Pusat Statistik. (2025). *Cerita Data Statistik untuk Indonesia: Transformasi Digital – Menelusuri Asosiasi Digitalisasi dengan Pendidikan dan Kesehatan*. Badan Pusat Statistik. <https://www.bps.go.id/id/publication/2025/04/30/675c728bd4ed47751a4f3c36/cerita-data-statistik-untuk-indonesia---transformasi-digital--menelusuri-asosiasi-digitalisasi-dengan-pendidikan-dan-kesehatan.html>

- Badri, M., Haerani, E., Syafria, F., Okfalisa, O., & Oktavia, L. (2025). Penerapan *Support Vector Machine* Dengan Smote Untuk Klasifikasi Sentimen Pada Data Ulasan Aplikasi Trading View. *Bulletin of Computer Science Research*, 6(1), 314–324. <https://doi.org/10.47065/bulletincsr.v6i1.793>
- Databokss. (2024). Indonesia peringkat ke-3 global dalam pemanfaatan aplikasi kesehatan. *Katadata Databoks*. <https://databoks.katadata.co.id/produk-konsumen/statistik/e7e5f757216d534/indonesia-peringkat-ke-3-global-memanfaatkan-aplikasi-kesehatan>
- Du, K.-L., Jiang, B., Lu, J., Hua, J., & Swamy, M. N. S. (2024). Exploring Kernel Machines and *Support Vector Machines*: Principles, Techniques, and Future Directions. *Mathematics*, 12(24), 3935. <https://doi.org/10.3390/math12243935>
- Egger, R., & Yu, J. (2022). A *Topic Modeling* Comparison Between LDA, NMF, Top2Vec, and *BERTopic* to Demystify Twitter Posts. *Frontiers in Sociology*, 7. <https://doi.org/10.3389/fsoc.2022.886498>
- Girsang, R. A. R. B. (2024). *ANALISIS SENTIMEN ULASAN PENGGUNA ALODOKTER DAN HALODOC MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)*. <http://repository.unas.ac.id/id/eprint/11876>
- Hua, Y. C., Denny, P., Wicker, J., & Taskova, K. (2024). A systematic review of *aspect-based sentiment analysis*: domains, methods, and trends. *Artificial Intelligence Review*, 57(11), 296. <https://doi.org/10.1007/s10462-024-10906-z>
- Imaduddin, H., A'la, F. Y., & Nugroho, Y. S. (2023). Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach. *International Journal of Advanced Computer Science and Applications*, 14(8). <https://doi.org/10.14569/IJACSA.2023.0140813>
- Indonesia Sentinel. (2025). *Survey shows 69% of young Indonesians rely on telehealth apps for healthcare needs*. Indonesia Sentinel. <https://indonesiasentinel.com/survey-shows-69-of-young-indonesians-rely-on-telehealth-apps-for-healthcare-needs/>
- Krippendorff, K. (2011). Computing Krippendorff's Alpha-Reliability. *Penn University of Pennsylvania*. <https://repository.upenn.edu/handle/20.500.14332/2089>
- Lia, A., Rahim, A., & Yoga Siswa, T. A. (2025). ANALISIS SENTIMEN APLIKASI MYSILOAM MENGGUNAKAN METODE NAÏVE BAYES. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1). <https://doi.org/10.23960/jitet.v13i1.5997>
- Ma, L., Chen, R., Ge, W., Rogers, P., Lyn-Cook, B., Hong, H., Tong, W., Wu, N., & Zou, W. (2025). AI-powered *topic modeling*: comparing LDA and *BERTopic* in analyzing opioid-related cardiovascular risks in women. *Experimental Biology and Medicine*, 250. <https://doi.org/10.3389/ebm.2025.10389>
- Muhajir, M., Rosadi, D., & Danardono. (2024). Integrating Decision Tree and BIRCH Clustering Algorithms of *BERTopic* for Analyzing Public Sentiment on Dirtyvote Movie. *Mathematical Modelling of Engineering Problems*, 11(12), 3391–3401. <https://doi.org/10.18280/mmep.111217>

- PT Siloam International Hospitals Tbk. (2025). *Annual Report 2025. Siloam Hospitals.* <http://www.siloamhospitals.com>
- Ren, H., Liu, Y., Naren, G., & Lu, J. (2024). The impact of multidirectional text typography on text readability in word clouds. *Displays*, 83, 102724. <https://doi.org/10.1016/j.displa.2024.102724>
- Sari Pranasti, A. (2025). Penerapan Analisis Sentimen Berbasis Aspek Menggunakan Mesin Vektor Pendukung pada Aplikasi Layanan Medis Digital. *Journal of Comprehensive Science (JCS)*, 4(2), 780–797. <https://doi.org/10.59188/jcs.v4i2.3047>
- Setiono, M. H. (2022). A KOMPARASI ALGORITMA DECISION TREE, RANDOM FOREST, SVM DAN K-NN DALAM KLASIFIKASI KEPUASAN PENUMPANG MASKAPAI PENERBANGAN. *INTI Nusa Mandiri*, 17(1), 32–39. <https://doi.org/10.33480/inti.v17i1.3420>
- Shin, Y., Darbon, J., & Karniadakis, G. E. (2023). Accelerating gradient descent and Adam via fractional gradients. *Neural Networks*, 161, 185–201. <https://doi.org/10.1016/j.neunet.2023.01.002>
- Wibowo, A., Indarti, I., & Laraswati, D. (2024). Komparasi Algoritma Decision Tree, Random Forest dan SVM untuk Prognosis COVID-19. *IMTechno: Journal of Industrial Management and Technology*, 5(2), 10–15. <https://doi.org/10.31294/imtechno.v5i2.2868>