

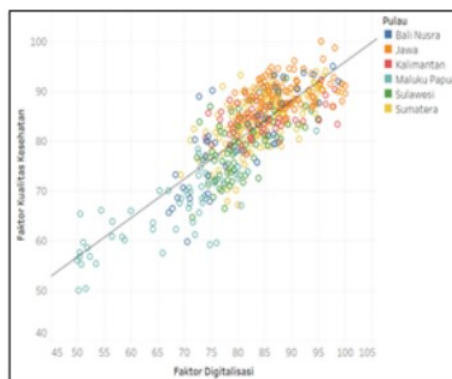
BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan era digital telah membawa perubahan signifikan dalam berbagai aspek kehidupan manusia, termasuk bidang kesehatan. Digitalisasi layanan kesehatan kini menjadi salah satu prioritas nasional untuk meningkatkan akses dan efisiensi pelayanan medis. Menurut data dari Katadata Media Network tahun 2020, sekitar 57% masyarakat Indonesia telah menggunakan aplikasi kesehatan digital, sehingga menempatkan Indonesia sebagai pasar terbesar ketiga secara global dalam pemanfaatan aplikasi kesehatan [1].

Selain itu, laporan Badan Pusat Statistik (BPS) dalam *Cerita Data Statistik untuk Indonesia Volume 2, Nomor 4 tahun 2025* menunjukkan bahwa peningkatan digitalisasi memiliki hubungan positif terhadap kualitas kesehatan. Berdasarkan hasil regresi yang ditampilkan pada gambar 1.1, setiap kenaikan 1 persen dalam tingkat digitalisasi berpotensi meningkatkan faktor kualitas kesehatan sebesar 0,164 poin dengan asumsi variabel lain tetap. Temuan ini menegaskan bahwa transformasi digital menjadi elemen penting dalam pembangunan sosial-ekonomi, terutama dalam meningkatkan efisiensi dan kualitas layanan kesehatan masyarakat Indonesia [2].



Gambar 1.1. Data Transformasi Digital BPS [2]

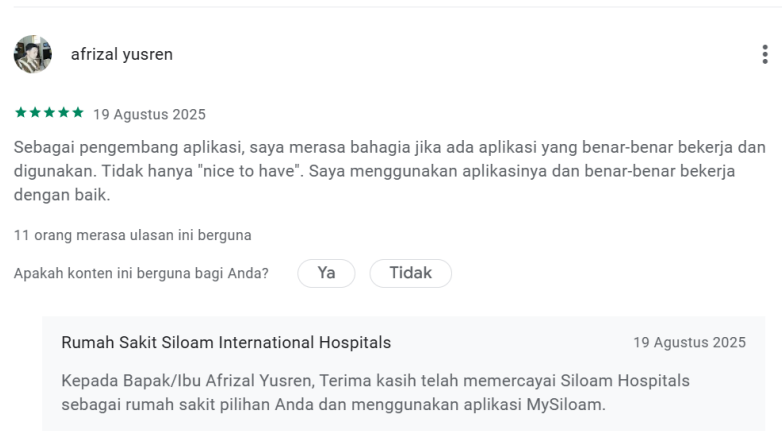
Sejalan dengan hal tersebut, pemerintah mendorong digitalisasi sistem kesehatan primer, antara lain penerapan rekam medis elektronik dan sistem informasi terintegrasi, guna mempercepat akses serta meningkatkan kualitas layanan kesehatan masyarakat [3]. Komitmen pemerintah terhadap transformasi digital di bidang kesehatan diperkuat dengan aplikasi *SATUSEHAT*. Hingga 15 Februari 2024, tercatat 8.362 fasilitas kesehatan telah terkoneksi dengan aplikasi ini dari total target 56.093 fasilitas kesehatan di seluruh Indonesia [4]. Di sisi lain, perubahan perilaku masyarakat juga semakin terlihat, khususnya di kalangan generasi muda. berdasarkan survei yang dipublikasikan oleh Indonesia Sentinel tahun 2023, sekitar 69% Generasi Z dan Milenial di Indonesia telah memanfaatkan aplikasi telehealth untuk kebutuhan layanan kesehatan sehari-hari [5]. Fakta-fakta tersebut menunjukkan bahwa digitalisasi layanan kesehatan berkembang pesat di Indonesia, baik dari sisi penyedia layanan maupun dari sisi pengguna.

Perkembangan tersebut turut mendorong rumah sakit besar di Indonesia untuk beradaptasi melalui inovasi digital. Salah satu contohnya adalah Siloam Hospitals Group, jaringan rumah sakit swasta terbesar di Indonesia dengan 41 rumah sakit dan kapasitas lebih dari 7.900 tempat tidur. Pada kuartal 1 tahun 2024, Siloam telah melayani lebih dari 1 juta pasien di seluruh Indonesia [6]. Sebagai bagian dari transformasi digital, Siloam meluncurkan aplikasi *MySiloam* yang menyediakan berbagai fitur antara lain pendaftaran online, pengecekan jadwal dokter, akses rekam medis elektronik, dan pembayaran digital. Hingga tahun 2025, aplikasi ini telah di unduh lebih dari satu juta kali dan memperoleh lebih dari 33 ribu ulasan pengguna di Google Play Store. Jumlah interaksi yang besar ini menunjukkan tingginya penerimaan Masyarakat, sekaligus membuka peluang analisis lebih lanjut terhadap persepsi pengguna.

Meskipun aplikasi *MySiloam* telah diunduh jutaan kali dan menerima puluhan ribu ulasan dari pengguna, informasi yang terkandung dalam ulasan tersebut belum dimanfaatkan secara optimal. Jumlah ulasan yang sangat besar membuat analisis manual menjadi sulit dilakukan secara efisien. Sebagai contoh, penelitian yang dilakukan oleh Lia et al. tahun 2025 menganalisis 1.995 ulasan pengguna *MySiloam*

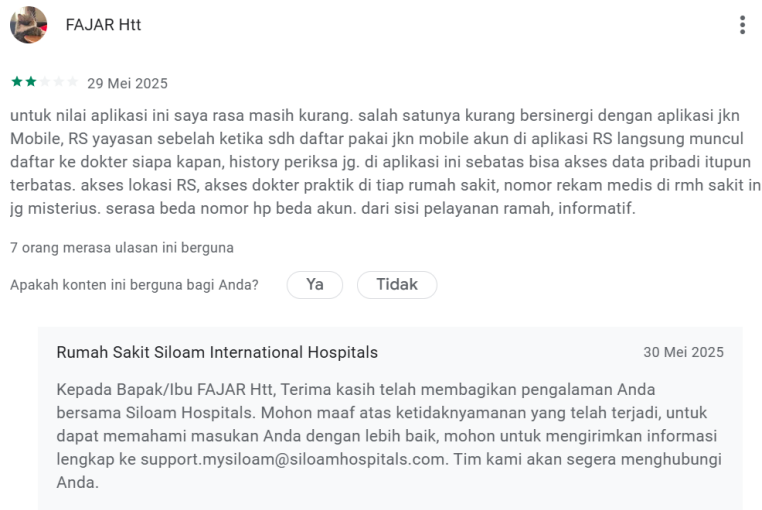
menggunakan metode *Naive Bayes* dengan representasi TF-IDF, dan menghasilkan akurasi sekitar 86% [7]. Penelitian ini relevan karena menjadikan *MySiloam* sebagai objek penelitian, namun analisis yang dilakukan masih terbatas pada klasifikasi sentimen sederhana (positif, negatif, dan netral) tanpa mengidentifikasi aspek-aspek spesifik dari aplikasi.

Beberapa ulasan pengguna di Google Play Store juga menunjukkan variasi persepsi yang signifikan terhadap aplikasi ini. Misalnya, salah satu pengguna menuliskan ulasan positif seperti pada gambar 1.3, dengan menyebut bahwa “*aplikasi ini benar-benar bekerja dan berguna, tidak hanya ‘nice to have’*”.



Gambar 1.2. Ulasan Positif Google Play Store Aplikasi MySiloam [8]

Namun, terdapat pula ulasan lain yang menyoroti keterbatasan fitur dan integrasi sistem, seperti pada komentar yang menyatakan bahwa “*aplikasi kurang bersinergi dengan JKN Mobile, akses dokter praktik di tiap rumah sakit terbatas, dan nomor rekam medis berbeda di tiap akun*”, dapat dilihat pada gambar 1.4.



Gambar 1.3. Ulasan Negatif Google Play Store Aplikasi MySiloam [8]

Variasi opini tersebut memperlihatkan bahwa persepsi pengguna terhadap *MySiloam* mencakup beragam aspek, mulai dari kemudahan penggunaan hingga kinerja sistem dan integrasi antarplatform. Oleh karena itu, diperlukan pendekatan analisis yang lebih mendalam untuk mengungkap persepsi pengguna berdasarkan aspek-aspek layanan yang relevan.

Keterbatasan tersebut menunjukkan bahwa meskipun analisis sentimen sederhana mampu memberikan gambaran umum tentang persepsi pengguna, metode tersebut belum cukup untuk menggali aspek-aspek spesifik yang memengaruhi pengalaman pengguna aplikasi kesehatan. Untuk mengatasi hal tersebut, digunakan pendekatan *Aspect Based Sentiment Analysis (ABSA)* yang merupakan jenis analisis sentiment *fine-grained* untuk mengidentifikasi aspek dan opini terkait dalam sebuah teks, sehingga memberikan wawasan yang lebih terfokus dibanding analisis sentimen tradisional yang hanya menghasilkan polaritas umum [9]. Misalnya, penelitian oleh Sari Pranasti, A., & Munggaran, L. C, tahun 2025, terkait aplikasi layanan medis digital di Indonesia menerapkan ABSA dengan SVM dan menemukan bahwa aspek-aspek seperti performa aplikasi, kualitas layanan, fitur, dan harga secara spesifik memengaruhi persepsi pengguna [10].

Dalam penerapannya, Aspect-Based Sentiment Analysis (ABSA) sering dikombinasikan dengan metode topic modeling untuk mengekstraksi aspek secara otomatis dari data ulasan yang besar. Berbagai teknik telah digunakan, seperti Latent Dirichlet Allocation (LDA) yang memodelkan dokumen sebagai campuran topik [11], Non-negative Matrix Factorization (NMF) yang memecah matriks dokumen dan kata untuk menemukan pola kata terhadap topik [12], dan Latent Semantic Analysis (LSA) yang mengidentifikasi hubungan laten antar kata menggunakan dekomposisi nilai singular [13]. Kemudian, seiring berkembangnya teknologi pemrosesan bahasa alami, muncul metode modern seperti BERTopic, yaitu *topic modeling* berbasis *transformer embeddings* yang mampu menghasilkan topik lebih koheren dan kontekstual dibanding pendekatan klasik [14].

Penelitian ini menggunakan BERTopic sebagai metode utama dalam proses ekstraksi aspek. BERTopic merupakan metode *topic modeling* berbasis *transformer embeddings* yang mampu menghasilkan representasi semantik yang lebih kaya dibandingkan model tradisional seperti LDA. Metode ini menggabungkan *embedding* dari model transformer, reduksi dimensi dengan UMAP, serta *class-based TF-IDF (c-TF-IDF)* untuk membentuk representasi topik yang lebih koheren dan informatif. Seperti dijelaskan dalam studi berjudul *BERTopic: Neural Topic Modeling with Class-based TF-IDF Procedure* oleh Maarten Grootendorst, BERTopic terbukti lebih adaptif dalam menangani variasi semantik dan konteks kalimat dibandingkan model klasik seperti LDA maupun NMF [14].

Selain itu, penelitian dengan judul *AI-powered Topic Modeling: Comparing LDA and BERTopic in Analyzing Opioid-Related Cardiovascular Risks in Women* menunjukkan bahwa *BERTopic* menghasilkan topik dengan koherensi semantik yang lebih baik dalam domain medis karena memanfaatkan representasi konteks dari model transformer [15]. Studi lain oleh Bayode Ogunleye et al. (2023) juga membuktikan bahwa metode pendekatan modern berbasis *embedding* seperti *BERTopic* mampu mengatasi keterbatasan tersebut dengan hasil topik yang lebih koheren dan kontekstual [16].

Dalam penggunaan default, *BERTopic* sering menggunakan *HDBSCAN* sebagai model clustering, tetapi modularitas ini memungkinkan penggunaan alternatif seperti *K-Means* atau *BIRCH* [17]. Oleh karena itu, dalam penelitian ini, tiga algoritma clustering *HDBSCAN*, *K-Means*, dan *BIRCH* akan dibandingkan dalam pipeline *BERTopic* untuk menentukan metode yang paling efektif dalam menghasilkan topik dengan stabilitas dan koherensi tinggi.

Selanjutnya, untuk mengklasifikasikan sentimen pada setiap aspek yang telah diidentifikasi, penelitian ini menggunakan algoritma *SVM*. Algoritma ini dipilih karena mampu melakukan klasifikasi secara non-linear dengan struktur pohon yang mudah diinterpretasikan serta efektif dalam menangani data dengan banyak fitur kategorikal maupun numerik. Misalnya, penelitian *Komparasi Algoritma Decision Tree, Random Forest, dan SVM untuk Prognosis COVID-19* menunjukkan bahwa *SVM* memiliki performa akurasi tertinggi dibandingkan algoritma lainnya, menegaskan kekuatannya dalam melakukan klasifikasi berbasis pola dan atribut yang kompleks [18].

Melalui perbandingan performa model berdasarkan metrik seperti akurasi, presisi, recall, dan F1-score, penelitian ini bertujuan untuk menentukan algoritma paling optimal dalam mengklasifikasikan sentimen pengguna terhadap aspek-aspek layanan MySiloam yang diidentifikasi melalui kombinasi metode *BERTopic* dan *BIRCH*.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana hasil ekstraksi aspek ulasan pengguna aplikasi MySiloam menggunakan metode *BERTopic* dengan perbandingan algoritma klasterisasi *BIRCH*, *HDBSCAN*, dan *K-Means*.
2. Bagaimana kualitas topik yang dihasilkan dari proses *topic modeling* pada ulasan pengguna aplikasi MySiloam?

3. Bagaimana performa model machine learning yaitu *SVM* dalam mengklasifikasikan sentimen pengguna pada setiap aspek layanan aplikasi *MySiloam*?

1.3. Batasan Masalah

Adapun batasan masalah yang digunakan pada penelitian ini berdasarkan rumusan masalah di atas adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini berupa ulasan pengguna aplikasi *MySiloam* yang diambil dari Google Play Store dan Apps Store.
2. Data yang digunakan merupakan ulasan pada rentang tahun periode 2019 – 2025 berjumlah 2.657 ulasan.
3. Ulasan yang dianalisis terbatas pada teks berbahasa Indonesia yang relevan dengan konteks layanan aplikasi.

1.4. Tujuan Penelitian

Berdasarkan uraian yang telah dipaparkan, tujuan yang hendak dicapai dalam penelitian ini adalah sebagai berikut:

1. Untuk mengetahui hasil ekstraksi aspek ulasan pengguna aplikasi *MySiloam* menggunakan metode *BERTopic* dengan perbandingan algoritma klusterisasi *BIRCH*, *HDBSCAN*, *daK-Means*.
2. Untuk mengevaluasi kualitas topik yang dihasilkan dari proses *topic modeling* pada ulasan pengguna aplikasi *MySiloam*.
3. Untuk menganalisis dan membandingkan performa model machine learning *Support Vector Machine* mengklasifikasikan sentimen pengguna pada setiap aspek layanan aplikasi *MySiloam*.

1.5. Manfaat Penelitian

Ada beberapa manfaat dari penelitian ini diantaranya sebagai berikut:

1. Memberikan kontribusi ilmiah terhadap pengembangan penelitian di bidang *Natural Language Processing* (NLP), khususnya melalui penerapan *Aspect-Based Sentiment Analysis* (ABSA) yang dikombinasikan dengan *BERTopic* untuk ekstraksi aspek secara otomatis.
2. Memberikan wawasan berbasis data bagi Siloam Hospitals Group dalam memahami persepsi pengguna terhadap aplikasi *MySiloam*, terutama pada aspek layanan yang perlu ditingkatkan.

1.6. Sistematika Penulisan

Struktur penulisan skripsi ini akan membantu memberikan panduan dalam menyusun laporan agar tetap sesuai dan menjadi acuan untuk mencapai tujuan penulisan laporan skripsi sesuai dengan yang diharapkan. Langkah-langkah dalam proses penyusunan skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan gambaran umum mengenai penelitian yang dilakukan, meliputi latar belakang yang mendasari dilakukannya penelitian, perumusan masalah, batasan masalah yang ditetapkan, tujuan penelitian, manfaat penelitian, serta sistematika penulisan skripsi secara keseluruhan.

BAB II TINJAUAN PUSTAKA

Bab ini memaparkan teori-teori dan penelitian terdahulu yang relevan sebagai dasar konseptual dalam penelitian ini. Materi yang dibahas mencakup *Google Play Store*, *MySiloam*, *Natural Language Processing* (NLP), *Text Mining*, *Data Preprocessing*, *Analisis Sentimen*, *Aspect-Based Sentiment Analysis* (ABSA), *BERTopic*, *Metrik Evaluasi Topic Modeling*, *SVM*, serta metode evaluasi model seperti *confusion matrix*, *accuracy*, *precision*, *recall*, *F1-score* dan *K-Fold Cross Validation*.

BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang tahapan ataupun metode penelitian yang akan digunakan. Tahapan tersebut yaitu studi literatur, pengumpulan data, eksplorasi data, praproses data, pemodelan topik, evaluasi topik, pelabelan, pembagian data, perancangan model, penerapan model, dan validasi model

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari setiap tahapan penelitian secara terperinci, termasuk seluruh proses pengolahan data, performa dari model klasifikasi yang diuji, serta evaluasi dan implementasi sistem berbasis web.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran-saran untuk penelitian selanjutnya agar dapat dikembangkan lebih lanjut. Kesimpulan disusun berdasarkan temuan empiris dan hasil analisis dari penelitian ini.

DAFTAR PUSTAKA

Bagian ini memuat seluruh referensi berupa jurnal, buku, laporan resmi, dan sumber daring yang digunakan dalam penyusunan skripsi ini, ditulis sesuai dengan format penulisan ilmiah yang berlaku.

LAMPIRAN

Dalam lampiran berisi informasi tambahan berupa data penting seperti surat izin penelitian.