

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital telah menghasilkan transformasi yang signifikan dalam layanan kesehatan, terutama melalui penerapan layanan *telemedicine*. Aplikasi *telemedicine* memfasilitasi penyediaan layanan konsultasi kesehatan kepada masyarakat tanpa harus datang langsung ke rumah sakit atau klinik, sehingga sangat membantu terutama ketika akses layanan tatap muka terbatas (Andi Sri Adinda, 2025). Di Indonesia, salah satu platform *telemedicine* yang paling umum adalah Halodoc, yang menawarkan layanan konsultasi medis secara *online*, mencakup berbagai layanan mulai dari pemesanan obat hingga pengujian laboratorium (Tarida Grace Wahyuni Margaretha Sidabutar, 2025). Popularitas Halodoc terlihat dari jumlah unduhan aplikasi di Google Play Store yang mencapai 10 juta kali, memperoleh rating rata-rata 4,8 dari 5 bintang, serta mendapat lebih dari 480 ribu lebih ulasan pengguna (Halodoc, 2025).

Ulasan pengguna pada *platform* aplikasi menjadi sumber informasi penting untuk mengetahui tingkat kepuasan dan pengalaman pengguna terhadap layanan yang diberikan. Sebagian besar ulasan menunjukkan penilaian positif, seperti kemudahan penggunaan aplikasi, ramah pengguna, dan bermanfaat bagi masyarakat dalam mengakses layanan kesehatan jarak jauh. Namun, tidak sedikit pula ulasan negatif yang muncul, terutama mengenai kesulitan dengan proses login, keterlambatan pengiriman obat, atau kendala teknis selama konsultasi video (Suri Wulandari, Medyantiwi Rahmawita Munzir, Nesdi Evrilyan Rozanda, 2024). Ulasan ini berfungsi sebagai sumber informasi penting untuk menilai kepuasan pengguna dan keluhan mengenai kualitas layanan, sehingga memungkinkan evaluasi melalui pendekatan analisis sentimen. Analisis sentimen ulasan yang berkaitan dengan aplikasi kesehatan telah terbukti secara efektif menggambarkan aspek positif dan negatif layanan secara objektif (Nuzuliarini Nuris, Eka Rini Yulia, 2021).

Dalam proses analisis sentimen mengenai ulasan pengguna Halodoc, beberapa tantangan memerlukan pertimbangan yang harus diperhatikan. Tahap awal prapemrosesan data seringkali tidak sederhana, karena ulasan yang dihasilkan pengguna bersifat tidak terstruktur dan penuh dengan emotikon, *hyperlink*, tagar, dan kesalahan ketik, sehingga mempersulit proses normalisasi data (Rahmadya Trias Handayanto, Herlawati, Prima Dina Atika & Ajif Yunizar Pratama Yusuf, 2021). Setelah data melalui tahap pembersihan, tantangan selanjutnya berkaitan dengan representasi fitur teks. Penerapan metodologi seperti TF-IDF sering menghasilkan jumlah fitur yang sangat besar, yang sebagian besar terlalu umum dan kurang relevan untuk membedakan sentimen. Kondisi ini mengarah pada dimensi data yang berlebihan, yang akibatnya mempengaruhi performa algoritma klasifikasi yang digunakan (Edwar et al., 2023). Jika tidak dilakukan seleksi fitur, model seperti *Support Vector Machine* (SVM) harus bekerja dengan dimensi data yang besar, sehingga menghasilkan waktu komputasi yang lebih lama serta peningkatan risiko *overfitting*.

Selain itu, distribusi kelas ulasan sering ditandai dengan ketidakseimbangan, misalnya, ada jumlah ulasan positif yang jauh lebih besar dibandingkan dengan ulasan negatif, membuat model kurang responsif terhadap keluhan pengguna yang sebenarnya penting untuk evaluasi. Untuk mengatasi ketidakseimbangan tersebut, dapat digunakan metode SMOTE (*Synthetic Minority Oversampling Technique*) yang berfungsi menyeimbangkan data dengan menambahkan sampel sintetis pada kelas minoritas (Hidayah et al., 2025).

Permasalahan utama dari penelitian ini adalah banyaknya fitur yang tidak relevan ada dalam evaluasi pengguna Halodoc, yang dapat mempengaruhi performa model klasifikasi. Kata-kata umum yang sering digunakan seperti “aplikasi” atau “dokter” sering kali muncul pada hampir semua ulasan, sehingga tidak efektif untuk membedakan polaritas sentimen. Kondisi ini menimbulkan masalah dimensi dalam algoritma *Support Vector Machine* (SVM), yang mengakibatkan pada waktu komputasi yang menjadi lebih lama dan peningkatan kerentanan terhadap *overfitting*. Oleh karena itu, penerapan seleksi fitur sangat penting untuk menghilangkan fitur *non-informatif* dan hanya mempertahankan kata-kata yang benar-benar relevan, yaitu kata-kata yang

mampu membedakan sentimen positif, negatif, dan netral, seperti “cepat”, “mudah”, dan “memuaskan” menunjukkan sentiment positif, sedangkan “lambat”, “error”, dan “rumit” menunjukkan sentiment negatif, serta kata-kata netral seperti “aplikasi”, “layanan”, “akun”, dan “proses”. Penelitian sebelumnya telah membuktikan bahwa penggabungan pemilihan fitur pada SVM dapat meningkatkan akurasi secara signifikan, seperti pada analisis sentimen Komisi Pemilihan Umum, yang menunjukkan peningkatan akurasi dari sekitar 66,49% menjadi 81,18% setelah penerapan seleksi fitur (Santoso et al., 2021). Hasil serupa juga terlihat pada penelitian mengenai evaluasi aplikasi Jamsostek Mobile, di mana SVM terbukti efektif dalam mengkategorikan sentimen pengguna, namun tetap memerlukan seleksi fitur untuk mengatasi jumlah fitur yang besar (Vina Fitriyana, Lutfi Hakim, Dian Candra Rini Novitasari & Asyhar, 2023).

Berbagai penelitian telah menunjukkan bahwa penerapan metodologi seleksi fitur *Information Gain* dapat meningkatkan performa klasifikasi sentimen ketika digunakan bersama dengan algoritma *Support Vector Machine* (SVM). Salah satu studi melakukan analisis perbandingan tentang performa *Information Gain* dengan *Chi-Squares* dalam konteks analisis sentimen ulasan aplikasi keuangan. Hasilnya menunjukkan bahwa kedua metode mampu mengurangi kompleksitas data sekaligus meningkatkan akurasi dan mempercepat waktu pelatihan model (Anelta Tirta Putri Subandono, 2025). Penelitian lain yang menganalisis ulasan aplikasi PeduliLindungi, membuktikan bahwa penerapan *Information Gain* pada SVM menghasilkan peningkatan akurasi yang signifikan, meningkat dari sekitar 74,33% menjadi 84% (Abror, 2023). Hasil ini menguatkan gagasan bahwa penggabungan SVM dan *Information Gain* merupakan solusi yang tepat untuk mengatasi tantangan yang ditimbulkan oleh dimensi fitur tinggi.

Meskipun demikian, beberapa penelitian juga menunjukkan bahwa penerapan seleksi fitur tidak selalu menghasilkan peningkatan performa yang signifikan pada semua kasus. Sebuah studi pada klasifikasi penyakit jantung menentukan bahwa meskipun *Information Gain* mampu mengurangi jumlah fitur, peningkatan akurasi tidak selalu terjadi pada setiap algoritma yang diuji (Siska Narulita, 2024). Selain itu, pada

analisis sentimen dalam tweet mengenai penanganan COVID-19, penerapan *Information Gain* berhasil mengurangi sebagian *noise* dari datase. Namun, itu belum sepenuhnya menyelesaikan masalah variasi bahasa (seperti slang, emotikon, dan kesalahan ketik) atau ketidakseimbangan kelas, menghasilkan kinerja yang kurang optimal dalam kelas minoritas (Fajri et al., 2023).

Berdasarkan berbagai penelitian tersebut, dapat diketahui bahwa kombinasi *Support Vector Machines* (SVM) dan *Information Gain* memiliki potensi dalam meningkatkan performa analisis sentimen. Namun, Sebagian besar penelitian tersebut masih berfokus pada analisis sentimen pada bidang lain seperti aplikasi keuangan, layanan pemerintahan, maupun media social. Penelitian yang secara khusus menganalisis ulasan pengguna aplikasi *telemedicine*, khususnya Halodoc masih terbatas. Selain itu, belum banyak penelitian yang secara eksplisit membandingkan performa model SVM sebelum dan sesudah penerapan seleksi fitur *Information Gain* pada dataset ulasan pengguna Halodoc. Oleh karena itu, diperlukan penelitian yang secara khusus menerapkan kombinasi metode SVM dan seleksi fitur *Information Gain* pada dataset ulasan aplikasi Halodoc.

Penelitian ini bertujuan untuk membangun model analisis sentimen yang mampu mengklasifikasikan ulasan pengguna Halodoc ke dalam kategori sentimen positif, negatif, dan netral menggunakan metode algoritma *Support Vector Machine* dengan seleksi fitur *Information Gain*. Penerapan seleksi fitur ini diharapkan dapat mengurangi jumlah fitur yang tidak relevan sehingga model menjadi lebih efisien dalam proses klasifikasi. Selain itu, penelitian ini juga diharapkan mampu memberikan gambaran mengenai penilaian pengguna terhadap layanan Halodoc serta menjadi bahan evaluasi bagi pengembang aplikasi dalam meningkatkan kualitas layanan kesehatan digital.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Ulasan pengguna Halodoc di Google Play Store jumlahnya sangat besar dan beragam, sehingga analisis secara manual tidak efisien. Banyaknya fitur atau kata juga dapat menurunkan akurasi model klasifikasi. Oleh karena itu, analisis sentimen menggunakan metode *Support Vector Machine* (SVM) dengan seleksi

fitur *Information Gain* dilakukan untuk mengetahui pengaruh terhadap performa model klasifikasi.

1.2.2 Metode Usulan

Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan penggunaan algoritma *Support Vector Machine* (SVM) sebagai metode klasifikasi yang dikombinasikan dengan *Information Gain* sebagai metode seleksi fitur. *Information Gain* berfungsi menyaring kata-kata yang relevan sehingga dimensi data berkurang. Selain itu, untuk membandingkan performa model SVM sebelum dan sesudah penerapan *Information Gain* untuk mengetahui pengaruh terhadap hasil klasifikasi sentimen.

1.2.3 Pertanyaan Penelitian

Berdasarkan rumusan masalah dan metode usulan, pertanyaan yang berkaitan dengan penelitian sebagai berikut:

1. Bagaimana performa algoritma SVM dalam mengklasifikasikan sentimen ulasan pengguna Halodoc?
2. Bagaimana pengaruh seleksi fitur *Information Gain* terhadap akurasi dan efisiensi model SVM dalam analisis sentimen ulasan pengguna Halodoc?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

1. Mengevaluasi performa algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen ulasan pengguna Halodoc.
2. Menganalisis pengaruh seleksi fitur *Information Gain* terhadap performa model *Support Vector Machine* (SVM) dalam analisis sentimen ulasan pengguna Halodoc.

1.3.2 Manfaat Penelitian

1. Bagi Halodoc: Memberikan masukan berdasarkan hasil analisis sentimen ulasan pengguna untuk meningkatkan kualitas layanan dan pengalaman pengguna.

2. Bagi peneliti dan Akademisi: Menjadi referensi dalam pengembangan metode analisis sentimen dengan penerapan SVM dan seleksi fitur *Information Gain*.
3. Bagi Pengguna: Memperoleh layanan yang lebih baik melalui perbaikan aplikasi berdasarkan masukan dari hasil analisis sentimen.

1.4 Batasan Masalah

Batasan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini terbatas pada ulasan pengguna Halodoc yang tersedia di Google Play Store, sehingga belum mencakup sumber ulasan lainnya.
2. Metode yang digunakan terbatas pada algoritma *Support Vector Machine* (SVM) untuk klasifikasi dan *Information Gain* untuk seleksi fitur.
3. Pada algoritma *Support Vector Machine* (SVM), penelitian ini hanya menggunakan kernel *linear*, *polynomial*, dan *Radial Basis Function* (RBF).
4. Penelitian ini hanya memproses ulasan Bahasa Indonesia, sehingga ulasan dalam bahasa lain tidak dianalisis.
5. Pada proses pelabelan, metode *Lexicon Based* hanya mempertimbangkan kata tunggal (*unigram*) tanpa memperhatikan konteks antar kata atau frasa.

1.5 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini mencakup latar belakang masalah, perumusan masalah, tujuan dan manfaat, penggambaran ruang lingkup masalah, serta struktur sistematis penulisan.

BAB II KAJIAN PUSTAKA

Bab ini membahas teori-teori yang mendasari penelitian, termasuk konsep analisis sentimen, algoritma *Support Vector Machine* (SVM), seleksi fitur dengan *Information Gain*, serta penelitian terdahulu yang relevan.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan penjelasan tentang tahapan penelitian, pengumpulan data, pra-pemrosesan, pemodelan, evaluasi kinerja model, serta skenario pengujian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil dan pembahasan berdasarkan analisis yang telah dilakukan.

BAB V PENUTUP

Bab ini memberikan Kesimpulan berdasarkan hasil analisis yang dilakukan serta saran untuk pengembangan kedepannya.