

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang akan digunakan dalam dokumen skripsi ini.

1.1 Latar Belakang

Perkembangan teknologi digital telah mendorong inovasi di berbagai sektor, termasuk bidang kesehatan. Salah satu inovasi yang berkembang pesat di Indonesia adalah layanan kesehatan berbasis *telemedicine*. *Telemedicine* merupakan praktik pemberian layanan kesehatan secara virtual atau jarak jauh dengan memanfaatkan teknologi komunikasi dan informasi untuk menghubungkan pasien dengan penyedia atau tenaga kesehatan. Sebagai bagian dari transformasi digital di sektor kesehatan, *telemedicine* bertujuan memperluas akses pelayanan kesehatan dan menjadi solusi untuk mengatasi berbagai tantangan dalam pelayanan kesehatan di Indonesia (Chairani, 2023).

Aplikasi Halodoc sebagai salah satu aplikasi *telemedicine* terkemuka yang menawarkan berbagai layanan, seperti konsultasi dengan dokter melalui *video* atau *chat*, pemesanan obat yang terintegrasi dengan layanan pengiriman, hingga pemeriksaan laboratorium. Kehadiran aplikasi Halodoc memberikan kemudahan bagi masyarakat untuk mengakses layanan kesehatan dengan cepat dan efisien tanpa harus mendatangi fasilitas kesehatan (Tabaeian dkk., 2024). Sebagai salah satu aplikasi kesehatan dengan rating tertinggi, aplikasi Halodoc tersedia di Google Play Store, platform distribusi digital resmi untuk perangkat Android maupun IOS. Melalui fitur komentar, ulasan, dan *rating*, pengguna dapat memberikan penilaian terhadap aplikasi, yang dapat menjadi indikator kualitas layanan serta pengalaman pengguna dalam memanfaatkan fitur-fitur yang disediakan.

Data ulasan ini menjadi sumber informasi penting bagi pengembang aplikasi untuk memahami persepsi pengguna terhadap produk mereka. Ulasan positif sering kali berisi apresiasi dan saran, sedangkan ulasan negatif cenderung mencerminkan keluhan pengguna terkait layanan yang diberikan. Untuk memahami persepsi publik dengan lebih mendalam, diperlukan proses analisis sentimen terhadap ulasan-ulasan ini.

Analisis sentimen merupakan pendekatan dalam pemrosesan bahasa alami (*Natural Language Processing* atau NLP) yang bertujuan mengidentifikasi opini atau sentimen dalam teks. Dalam konteks ulasan aplikasi, analisis sentimen bertujuan menentukan apakah ulasan tersebut bersentimen positif, negatif, atau netral. Proses ini melibatkan tahapan-tahapan seperti *tokenization*, penghapusan kata yang tidak bermakna (*stop-words*), dan *lemmatization* untuk menghasilkan data yang lebih bersih dan relevan. Setelah proses *preprocessing* selesai, fitur-fitur yang relevan diekstraksi dan digunakan dalam metode klasifikasi (Anagha dkk., 2020).

Dalam melakukan analisis sentimen, ekstraksi fitur menjadi salah satu langkah penting yang dapat memengaruhi performa suatu metode. Ekstraksi ini melibatkan pemilihan fitur untuk mengubah teks menjadi representasi numerik yang dapat diklasifikasikan (Rusli dkk., 2020). Beberapa metode ekstraksi fitur yang paling umum digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Bag-of-Words* (BoW). TF-IDF merupakan salah satu teknik dalam ekstraksi fitur yang berfungsi sebagai penghitung bobot setiap kata dalam dokumen (Pratomo dkk., 2021). Sedangkan BoW berfungsi dengan cara menganalisis kata-kata dalam suatu dokumen, kemudian merepresentasikan dokumen tersebut dengan menghitung frekuensi kemunculan setiap kata di dalamnya (Deepu, Pethuru, & Rajaraajeswari, 2016).

Baik TF-IDF maupun BoW memiliki keunggulan masing-masing. TF-IDF efektif dalam menangkap bobot kata yang lebih spesifik, sementara BoW menawarkan kesederhanaan dan kecepatan dalam penggunaannya. Dalam berbagai penelitian, performa kedua metode ini sangat bergantung pada metode klasifikasi dan karakteristik data yang digunakan.

Penelitian oleh (Hendra, 2021) menggunakan *Naïve Bayes Classifier* (NBC) untuk analisis sentimen ulasan pengguna aplikasi Halodoc di Google Play Store. Proses penelitiannya meliputi tahapan *preprocessing* data, seperti *tokenization*, *lowercasing*, penghapusan *stop-words*, dan pembobotan kata menggunakan *TF-IDF*. Hasil penelitian menunjukkan bahwa NBC menghasilkan tingkat akurasi 81,68% dan nilai AUC (*Area Under Curve*) sebesar 0,756, yang tergolong klasifikasi *fair* (cukup baik).

Penelitian lain yang dilakukan oleh Farhan dkk. (2022) menggunakan metode *Support Vector Machine* (SVM) dengan model BoW untuk menganalisis sentimen terkait Pemilihan Kepala Daerah (Pilkada) 2020 yang berlangsung di tengah pandemi COVID-19. Dengan data sebanyak 400 cuitan Twitter (187 positif dan 213 negatif), hasil evaluasi menunjukkan tingkat akurasi, *precision*, *recall*, dan *F1-score* sebesar 87,5%, yang mengindikasikan kemampuan metode tersebut dalam mengklasifikasikan sentimen secara efektif. Selain itu, penelitian yang dilakukan oleh Pratomo dkk. (2021) mengkaji performa klasifikasi ulasan film dari situs IMDb dengan menggunakan kombinasi metode ekstraksi fitur *TF-IDF* dan leksikon *SentiWordNet*, serta fitur seleksi *Information Gain* (IG). *Dataset* yang digunakan terdiri atas 2.000 ulasan berbahasa Inggris. Hasil penelitian menunjukkan bahwa kombinasi *TF-IDF* dan *SentiWordNet* menghasilkan akurasi 73,31%, lebih rendah dibandingkan dengan penggunaan *TF-IDF* saja (81,04%). Namun, penerapan IG dengan *threshold* optimal mampu meningkatkan akurasi klasifikasi menggunakan *K-Nearest Neighbor* (KNN) hingga 81,04%, dibandingkan tanpa IG yang hanya mencapai akurasi 77,80%.

Penelitian lain oleh Putra dkk. (2023) membandingkan teknik ekstraksi fitur BoW dan *TF-IDF* dalam analisis sentimen menggunakan metode *Support Vector Machine* pada data teks Twitter. BoW menghitung frekuensi kemunculan setiap kata dalam dokumen dan merepresentasikannya sebagai vektor numerik, sementara *TF-IDF* menambahkan bobot pada kata-kata berdasarkan frekuensinya di seluruh dokumen. Hasil penelitian menunjukkan bahwa BoW unggul dalam *Recall* dengan nilai 89%, yang menandakan efektivitasnya dalam mendeteksi sentimen positif, sementara *TF-IDF* memberikan hasil lebih baik dalam *Accuracy* (86%), *Precision* (85%), dan *F1-Score* (85%). Sementara itu, penelitian oleh Gifari dkk. (2022) menggunakan metode *TF-IDF* sebagai pembobotan kata dan metode *Support Vector Machine* untuk analisis sentimen ulasan film. Hasil penelitian menunjukkan bahwa kombinasi *TF-IDF* dan metode *Support Vector Machine* mampu menghasilkan akurasi 85%, *precision* 100%, *recall* 70%, dan *F1-score* 82%. Penelitian ini menegaskan pentingnya proses *preprocessing* data guna meningkatkan performa analisis, terutama dalam menangani kata-kata tidak baku atau singkatan yang kerap dijumpai dalam ulasan berbasis teks.

Meskipun penelitian tentang analisis sentimen menggunakan *TF-IDF*, BoW, dan SVM telah banyak dilakukan, sebagian besar penelitian tersebut menggunakan domain yang berbeda, seperti ulasan film, Twitter, atau isu politik. Oleh karena itu, diperlukan eksplorasi

lebih mendalam untuk memahami bagaimana metode ekstraksi fitur ini bekerja pada ulasan aplikasi kesehatan khususnya Halodoc yang ditulis dalam bahasa Indonesia. Hal ini menjadi tantangan tersendiri karena teks ulasan sering mengandung kata-kata tidak baku, singkatan, atau istilah khusus yang membutuhkan proses *preprocessing* data yang baik.

Dalam konteks ini, metode *Support Vector Machine* (SVM) telah menjadi salah satu teknik yang sering digunakan dalam klasifikasi sentimen, memanfaatkan pendekatan *Structural Risk Minimization* (SRM). Pendekatan ini bertujuan untuk menemukan *hyperplane* terbaik yang dapat memisahkan kelas data secara optimal. Selain itu, tingkat akurasi hasil klasifikasi dengan metode SVM sangat bergantung pada fungsi kernel dan parameter yang digunakan (Parapat & Furqon, 2018).

Dengan demikian, penelitian ini bertujuan membandingkan performa metode ekstraksi fitur TF-IDF dan BoW menggunakan metode *Support Vector Machine* pada *dataset* ulasan pengguna Halodoc di Google Play Store. Berdasarkan uraian latar belakang yang telah dipaparkan, penelitian ini akan menguji bagaimana masing-masing metode ekstraksi fitur tersebut bekerja ketika diproses menggunakan metode *Support Vector Machine*. Oleh karena itu, dirancang sebuah penelitian dengan judul Perbandingan Ekstraksi Fitur TF-IDF dan BoW pada Analisis Sentimen Ulasan Pengguna Aplikasi Halodoc Menggunakan Metode *Support Vector Machine*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah bagaimana perbandingan performa metode ekstraksi fitur BoW dan TF-IDF dalam analisis sentimen terhadap ulasan pengguna aplikasi Halodoc menggunakan metode *Support Vector Machine* (SVM) berdasarkan metrik evaluasi akurasi, presisi, *recall*, dan *F1-score*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan penelitian ini adalah untuk membandingkan performa metode ekstraksi fitur BoW dan TF-IDF dalam mendukung sentimen terhadap ulasan pengguna aplikasi Halodoc. Perbandingan ini dilakukan dengan menggunakan metode *Support Vector Machine* (SVM) berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score*.

1.4 Manfaat Penelitian

Berdasarkan permasalahan yang telah diuraikan dan tujuan penelitian yang telah ditetapkan, penelitian ini diharapkan memberikan manfaat yang signifikan. Bagi perusahaan Halodoc, hasil penelitian ini dapat menjadi bahan evaluasi dalam upaya meningkatkan kualitas layanan mereka berdasarkan sentimen yang terkandung dalam ulasan pengguna. Sementara itu, bagi peneliti lain, penelitian ini memberikan pemahaman yang lebih baik tentang performa yang diperoleh dari penggunaan metode ekstraksi data BoW dan TF-IDF pada algoritma SVM, sehingga dapat menjadi referensi untuk penelitian lebih lanjut di bidang serupa.

1.5 Ruang Lingkup

Ruang lingkup penelitian ini dibuat untuk memastikan skripsi berjalan terarah dan mencapai tujuan yang diharapkan. Berikut adalah ruang lingkupnya :

1. Data yang digunakan dalam penelitian ini merupakan data ulasan aplikasi Halodoc yang diperoleh dari *platform* Google Play Store. Pengambilan data dilakukan menggunakan metode *scraping* dengan bantuan library *google-play-scraper*. Data yang dikumpulkan sebanyak 10.994 ulasan. Proses *scraping* dilakukan pada tanggal 11 November 2024.
2. Klasifikasi sentimen *review* dilakukan sebanyak dua kategori yaitu positif dan negatif berdasarkan *rating* 1-5.
3. Penelitian ini hanya menggunakan ulasan dalam bahasa Indonesia yang tersedia di Google Play Store.
4. Proses implementasi dilakukan menggunakan bahasa pemrograman Python dengan perangkat lunak Google Colaborator.

1.6 Sistematika Penulisan

Sistematika penulisan ini memberikan gambaran mengenai isi laporan skripsi secara urut dan sistematis. Skripsi ini disusun dalam lima bab, dengan rincian sebagai berikut.

BAB I PENDAHULUAN

Bab ini menyajikan latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan skripsi perbandingan metode ekstraksi fitur

BoW dan TF-IDF dalam analisis sentimen terhadap ulasan pengguna aplikasi Halodoc menggunakan metode *Support Vector Machine (SVM)*.

BAB II LANDASAN TEORI

Bab ini menguraikan teori-teori yang mendasari penelitian, dimulai dengan konsep analisis sentimen sebagai proses memahami opini pengguna dari data teks. Penelitian ini menggunakan *Natural Language Processing (NLP)* dan *text mining* untuk pengolahan data teks melalui beberapa tahapan, yaitu *data cleaning*, normalisasi, tokenisasi, *stemming*, dan penghapusan *stopwords*. Data teks kemudian direpresentasikan dalam bentuk numerik menggunakan metode ekstraksi fitur TF-IDF dan *Bag of Words (BoW)*. Objek studi berupa ulasan pengguna aplikasi Halodoc dari Google Play Store. Model klasifikasi menggunakan metode *Support Vector Machine (SVM)*, dengan penjelasan prinsip kerja, fungsi kernel, dan parameter yang digunakan. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*, serta *confusion matrix* untuk memberikan gambaran yang lebih mendalam mengenai distribusi prediksi model, termasuk jumlah prediksi benar (*True Positive* dan *True Negative*) serta prediksi salah (*False Positive* dan *False Negative*). Penelitian ini menggunakan metodologi CRISP-DM dan memanfaatkan pustaka Python seperti Scikit-learn, NLTK, spaCy, Pandas, dan Matplotlib untuk mendukung proses analisis dan visualisasi data.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metodologi yang digunakan dalam skripsi perbandingan metode ekstraksi fitur *TF-IDF* dan *BoW* dalam analisis sentimen terhadap ulasan pengguna aplikasi Halodoc di Google Play Store dengan menggunakan metode *Support Vector Machine (SVM)*. Bab ini membahas alur penelitian, pengumpulan data melalui teknik *web scraping*, proses labelisasi data, prapemrosesan data, pembagian data serta proses klasifikasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil analisis sentimen pada ulasan pengguna aplikasi Halodoc dengan membandingkan performa metode ekstraksi fitur *Bag of Words (BoW)* dan

TF-IDF menggunakan metode *Support Vector Machine* (SVM). Selain itu, bab ini juga membahas secara mendalam hasil yang diperoleh dari penelitian ini.

BAB V PENUTUP

Bab ini menyajikan kesimpulan dan saran berdasarkan hasil penelitian mengenai perbandingan metode ekstraksi fitur BoW dan TF-IDF dalam analisis sentimen terhadap ulasan pengguna aplikasi Halodoc menggunakan metode *Support Vector Machine* (SVM).