

BAB I

PENDAHULUAN

1.1. Latar Belakang

Di era digital saat ini, teknologi informasi telah menjadi bagian yang tidak terpisahkan dari kehidupan sehari-hari. Teknologi memengaruhi berbagai aspek, termasuk komunikasi, pendidikan, transportasi, hingga gaya hidup masyarakat [1]. Salah satu dampak signifikan dari perkembangan ini adalah kemudahan dalam mengakses dan berbagi informasi, termasuk melalui ulasan atau review pengguna terhadap layanan yang mereka gunakan. Ulasan pengguna kini menjadi salah satu sumber informasi utama dalam menilai kualitas suatu layanan atau produk, termasuk layanan transportasi daring seperti Maxim, Grab, dan Gojek [2].

Platform seperti *Google Play Store* dan *App Store* memungkinkan pengguna untuk meninggalkan komentar dan penilaian terkait pengalaman mereka saat menggunakan aplikasi transportasi. Ulasan ini umumnya memuat pendapat mengenai berbagai aspek seperti kenyamanan, kecepatan layanan, harga, kemudahan aplikasi, hingga sikap pengemudi. Dalam satu ulasan, pengguna bisa menyampaikan beberapa opini sekaligus, yang mencerminkan beragam aspek dengan sentimen berbeda-beda. Sebagai contoh, pengguna mungkin menulis: "Aplikasi mudah digunakan, tapi pengemudinya sering terlambat." Kalimat tersebut menunjukkan aspek antarmuka aplikasi dengan sentimen positif, namun juga mengandung kritik terhadap ketepatan waktu layanan dengan sentimen negatif [3].

Keberagaman isi ulasan dan banyaknya data yang tersedia menyebabkan analisis secara manual menjadi tidak efisien. Baik perusahaan penyedia layanan maupun calon pengguna akan kesulitan untuk memahami persepsi keseluruhan terhadap suatu layanan jika harus membaca ribuan ulasan secara satu per satu. Oleh karena itu, dibutuhkan pendekatan otomatis yang dapat mengekstraksi informasi penting dari ulasan secara efisien dan akurat.

Dalam hal ini, analisis sentimen berbasis aspek (*Aspect-Based Sentiment Analysis/ABSA*) hadir sebagai solusi. ABSA merupakan metode dalam pemrosesan bahasa alami (*Natural Language Processing/NLP*) yang mampu mengidentifikasi

aspek-aspek spesifik dalam sebuah ulasan dan menentukan sentimen yang terkait dengan masing-masing aspek tersebut [4]. Pendekatan ini lebih unggul dibandingkan analisis sentimen umum, karena mampu memberikan wawasan yang lebih mendalam mengenai pendapat pengguna terhadap elemen-elemen layanan tertentu.

Penerapan analisis sentimen telah banyak dikembangkan oleh berbagai peneliti. Dalam sejumlah studi sebelumnya, pendekatan analisis sentimen dilakukan menggunakan beragam teknik ekstraksi fitur teks, seperti TF-IDF dan Word Embedding (misalnya *Word2Vec* dan *FastText*). TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan metode statistik yang digunakan untuk menghitung bobot suatu kata dalam dokumen berdasarkan tingkat kemunculannya. Penelitian [5] mengungkapkan bahwa TF-IDF memiliki kelebihan dalam mengolah dataset berukuran kecil hingga menengah, dengan capaian akurasi sebesar 85% saat dipadukan dengan algoritma *Support Vector Machine* (SVM) dalam menganalisis sentimen pada ulasan produk di Indonesia. Sementara itu, studi [6] menunjukkan bahwa kombinasi SVM dan TF-IDF menghasilkan akurasi yang lebih baik dibandingkan dengan penggunaan *Word2Vec* dalam analisis sentimen pada platform *TikTok Shop*.

Penelitian selanjutnya menunjukkan bahwa analisis sentimen terhadap komentar YouTube dapat memberikan gambaran penerimaan publik terhadap suatu produk, seperti Samsung Galaxy Z Flip 3. Dengan menggunakan pendekatan CRISP-DM dan membandingkan beberapa metode klasifikasi, yaitu *Naïve Bayes*, SVM, dan K-NN, ditemukan bahwa metode *Support Vector Machine* (SVM) memberikan hasil paling akurat. SVM unggul karena mampu menangani data teks yang kompleks dan tidak seimbang dengan akurasi tinggi, mencapai rata-rata 96,43% di berbagai aspek, seperti desain, harga, spesifikasi, dan citra merek. Keunggulan SVM terletak pada kemampuannya memisahkan data secara optimal dan memberikan performa stabil pada analisis multi-aspek [7]. Penelitian selanjutnya membandingkan tiga algoritma klasifikasi *Support Vector Machine* (SVM), *Perceptron*, dan *Multinomial Naïve Bayes* (MNB) untuk analisis sentimen ulasan produk di Tokopedia, salah satu platform *e-commerce* terbesar di Indonesia. Dari 2.000 data ulasan yang dianalisis, *Multinomial Naïve Bayes*

menunjukkan performa terbaik dengan akurasi tertinggi sebesar 84% dan presisi 96%, mengungguli SVM dan Perceptron. Evaluasi melalui confusion matrix juga memperkuat keunggulan MNB dalam mengklasifikasikan sentimen positif dan negatif secara lebih akurat. Temuan ini menegaskan bahwa MNB merupakan algoritma yang paling efektif dalam konteks analisis sentimen e-commerce dan dapat dijadikan acuan dalam penelitian lanjutan [8].

Berdasarkan uraian dan permasalahan yang telah dijabarkan sebelumnya, penelitian ini berfokus pada Analisis Sentimen Berbasis Aspek terhadap ulasan pengguna Maxim, Grab, dan Gojek menggunakan *Support Vector Machine (SVM)* dengan menggunakan representasi fitur TF-IDF. SVM dipilih karena kemampuannya menangani data tidak linier dan hasil akurasi tinggi dalam analisis sentimen, sementara Untuk mendapatkan aspek-aspek yang berhubungan dengan aplikasi dipilih menggunakan topic modelling LDA karena LDA adalah metode pemodelan topik yang efektif untuk membuat konteks kumpulan dokumen dan memberikan topik yang lebih berkualitas. Topik ini kemungkinan besar akan dicakup oleh ulasan. Sehingga dapat mengukur sentimen seseorang terhadap layanan atau produk. Penelitian ini bertujuan memberikan solusi otomatis dalam mengolah ribuan ulasan berbahasa Indonesia yang mengandung karakteristik informal, sekaligus menghasilkan insight akurat bagi perusahaan untuk meningkatkan kualitas layanan berdasarkan aspek yang paling banyak dikomentari pengguna, sehingga diharapkan dapat menjadi referensi pengembangan sistem analisis sentimen di sektor transportasi daring maupun industri berbasis umpan balik pengguna lainnya.

1.2. Perumusan Masalah

1.2.1. Permasalahan

Ulasan pengguna pada aplikasi Maxim, Grab, dan Gojek di *Google Play Store* mengandung opini yang mencerminkan kepuasan terhadap berbagai aspek layanan. Namun, volume data yang besar dan beragam menyulitkan proses analisis manual. Oleh karena itu, diperlukan metode otomatis untuk mengklasifikasikan sentimen berbasis aspek menggunakan pendekatan yang tepat seperti LDA digunakan untuk mengekstraksi aspek-aspek utama dari ulasan

aplikasi maxim, grab dan gojek kemudian *Support Vector Machine* (SVM) untuk mengidentifikasi sentimen dari masing-masing aspek. Adapun permasalahan yang dirumuskan dalam penelitian ini meliputi:

1. Bagaimana metode Latent Dirichlet Allocation (LDA) dapat digunakan untuk mengidentifikasi aspek-aspek utama dalam ulasan pengguna Maxim, Grab, dan Gojek?
2. Bagaimana performa algoritma SVM dalam menganalisis sentimen berbasis aspek dari ulasan pengguna Maxim, Grab, dan Gojek?
3. Bagaimana pengaruh penerapan metode SMOTE terhadap performa algoritma SVM dalam menangani ketidakseimbangan data pada klasifikasi sentimen berbasis aspek?
4. Bagaimana akurasi, presisi, recall, dan F1-score dari model klasifikasi yang dibangun menggunakan TF-IDF dan SVM?

1.2.2. Solusi Permasalahan

Penelitian ini mengusulkan pendekatan dengan tahapan sebagai berikut:

1. LDA digunakan untuk mengekstraksi aspek-aspek utama dari ulasan pengguna Maxim, Grab dan Gojek secara otomatis.
2. Menggunakan metode TF-IDF untuk merepresentasikan teks ulasan sebagai fitur numerik yang dapat diproses oleh algoritma klasifikasi.
3. Melakukan klasifikasi sentimen pada masing-masing aspek menggunakan algoritma *Support Vector Machine* (SVM) untuk menentukan polaritas (positif atau negatif) dari setiap aspek ulasan.
4. Menerapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menyeimbangkan jumlah data pada kelas sentimen sebelum proses klasifikasi.
5. Melakukan pengujian model dengan variasi parameter dan pembagian data guna memperoleh performa klasifikasi yang optimal.
6. Mengevaluasi hasil klasifikasi berdasarkan metrik akurasi, presisi, recall, dan F1-score untuk menilai kinerja model terhadap data yang telah dianalisis.

1.2.3. Pertanyaan Penelitian

1. Bagaimana efektivitas kombinasi metode LDA dan performa algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen pada masing-masing aspek terhadap ulasan pengguna Maxim, Grab, dan Gojek dengan menggunakan representasi fitur TF-IDF?
2. Bagaimana pengaruh variasi parameter dalam TF-IDF terhadap akurasi, presisi, recall, dan F1-score dari model klasifikasi sentimen yang dibangun?
3. Bagaimana pengaruh penerapan metode *Synthetic Minority Over-sampling Technique* (SMOTE) terhadap kinerja algoritma SVM dalam meningkatkan performa klasifikasi pada data ulasan yang tidak seimbang?

1.3. Tujuan dan Manfaat

1.3.1. Tujuan Penelitian

1. Menganalisis sentimen berbasis aspek dengan kombinasi metode LDA, SMOTE dan SVM terhadap ulasan pengguna aplikasi Maxim, Grab, dan Gojek yang diperoleh dari *Google Play Store*.
2. Menerapkan algoritma *Support Vector Machine* (SVM) dengan representasi fitur TF-IDF untuk melakukan klasifikasi sentimen.
3. Mengevaluasi performa model klasifikasi berdasarkan metrik akurasi, presisi, recall, dan F1-score dengan mempertimbangkan pengaruh variasi parameter.

1.3.2. Manfaat Penelitian

1. Memberikan wawasan kepada penyedia layanan transportasi online seperti Maxim, Grab, dan Gojek mengenai persepsi pengguna terhadap layanan tersebut yang dapat berguna sebagai bahan evaluasi dalam peningkatan layanan dalam aspek-aspek yang dibahas.
2. Menyediakan metode analisis sentimen berbasis aspek yang efektif dan dapat diterapkan secara otomatis untuk mendukung pengambilan keputusan dalam peningkatan layanan.
3. Menjadi referensi bagi penelitian selanjutnya dalam bidang analisis sentimen dan klasifikasi teks menggunakan algoritma *Support Vector Machine* (SVM), SMOTE dan representasi fitur TF-IDF.

1.4. Batasan Masalah

Agar permasalahan dan penyusunan penelitian dapat dilakukan secara terarah dan tercapai sesuai dengan yang diharapkan, penelitian ini perlu mempertimbangkan beberapa batasan yang relevan. Dengan memahami dan mengikat batasan-batasan ini, penelitian dapat fokus pada hal-hal yang penting dan menghindari kerumitan yang tidak perlu.:

1. Ulasan layanan transportasi online diambil dari platform Maxim, Grab, dan Gojek .
2. Ulasan yang diolah adalah ulasan berbahasa Indonesia .
3. Metode TF-IDF digunakan dalam proses ekstraksi fitur teks .
4. Metode *Support Vector Machine* (SVM) digunakan untuk analisis sentimen berbasis aspek.
5. *Topic modeling* yang digunakan yaitu *Latent Dirichlet Allocation* (LDA)
6. Sentimen pada masing-masing aspek diklasifikasikan ke dalam dua kategori, yaitu positif dan negatif .
7. Untuk mengatasi jumlah data yang tidak seimbang antara kelas mayoritas dan minoritas, penelitian ini menggunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*).
8. Evaluasi model dilakukan berdasarkan metrik akurasi, presisi, recall, dan F1-score.

1.5. Sistematika Penulisan

Sistematika proposal ini disusun dalam tiga bab sebagai berikut:

BAB 1 PENDAHULUAN

Bagian ini memberikan tinjauan umum mengenai penelitian yang akan dilakukan. Isinya mencakup latar belakang, perumusan masalah, batasan-batasan yang ditetapkan, tujuan serta manfaat dari penelitian ini, serta sistematika dalam penyusunan laporan.

BAB 2 TINJAUAN PUSTAKA

Bagian ini menjelaskan mengenai teori-teori pendukung dari berbagai sumber yang berhubungan dengan penelitian ini dan penelitian terdahulu yang membahas topik serupa.

BAB 3 METODOLOGI PENELITIAN

Bagian ini menjelaskan tentang arsitektur sistem, proses pra-pemrosesan data, langkah-langkah atau sistematisasi dari setiap metode yang diusulkan, serta skenario pengujian yang akan dilakukan.

BAB 4 HASIL DAN PEMBAHASAN

Bagian ini berisi hasil dan pembahasan implementasi dari metode yang diusulkan dan analisis terhadap hasil implementasi

BAB 5 PENUTUP

Bagian ini memuat kesimpulan dari hasil penelitian yang dilakukan, serta saran untuk pengembangan pada penelitian selanjutnya berdasarkan hasil analisis yang di peroleh pada bab 4