

BAB I PENDAHULUAN

1.1 Latar Belakang

E-commerce berkembang seiring dengan kemajuan teknologi informasi dan meningkatnya penggunaan internet di Indonesia, yang mendorong pergeseran pola transaksi dari konvensional menjadi digital[1]. Melalui platform marketplace, konsumen kini dapat membeli berbagai produk secara daring dengan cara yang lebih praktis dan efisien[2]. Selain memudahkan transaksi, e-commerce juga memungkinkan konsumen untuk berbagi pengalaman melalui ulasan produk yang mereka beli[3]. Perkembangan ini turut mengubah kebiasaan masyarakat dalam berbelanja, di mana konsumen membutuhkan informasi tambahan sebagai bahan pertimbangan, mengingat mereka tidak dapat melihat atau mencoba produk secara langsung sebelum membeli[4].

Ulasan produk dari pengguna lain menjadi sumber informasi yang sangat berharga karena memberikan gambaran nyata mengenai pengalaman mereka dalam menggunakan produk. Melalui ulasan ini, calon pembeli dapat memahami kelebihan dan kekurangan suatu produk berdasarkan kebutuhan dan kondisi pribadi mereka masing-masing[5]. Dengan demikian, ulasan pelanggan secara online terbukti menjadi faktor utama yang memengaruhi persepsi dan kepercayaan konsumen dalam proses pengambilan keputusan pembelian[6].

Dalam industri kecantikan, terdapat beragam produk dari berbagai kategori dan merek yang terus berkembang mengikuti kebutuhan pasar. Setiap produk dirancang untuk memenuhi kebutuhan spesifik pengguna, baik untuk merawat wajah, tubuh, rambut, maupun aspek lainnya. Untuk membantu konsumen memilih produk yang sesuai, hadir platform komunitas seperti *Female Daily* yang menjadi tempat berbagi informasi dan pengalaman mengenai berbagai produk kecantikan. Melalui platform ini, pengguna dapat menuliskan ulasan berdasarkan pengalaman pribadi, sehingga konsumen lain bisa mempertimbangkan kecocokan produk tersebut dengan kondisi mereka masing-masing. Hal ini penting karena setiap produk umumnya memiliki manfaat yang berbeda-beda tergantung pada kebutuhan dan karakteristik individu yang menggunakannya[5].

Meskipun ulasan produk kecantikan sangat membantu konsumen dalam memilih produk yang tepat, seringkali ulasan tersebut berisi berbagai pengalaman yang sangat beragam. Tidak hanya membahas keseluruhan produk, tetapi juga menyoroti berbagai aspek spesifik seperti kualitas, kemasan, harga, dan yang lainnya. Keberagaman isi ulasan ini membuatnya menjadi sumber informasi yang kaya namun juga kompleks, sehingga sulit bagi konsumen untuk memahami mana aspek yang paling berpengaruh terhadap kepuasan atau ketidakpuasan mereka secara menyeluruh[7].

Kompleksitas ulasan yang memuat beragam pandangan pengguna membuat proses pemahaman terhadap informasi menjadi tidak sederhana. Dalam situasi ini, konsumen maupun produsen seringkali mengalami kesulitan dalam mengidentifikasi fokus ulasan yang sebenarnya, terutama ketika berbagai opini tercampur dalam satu ulasan panjang. Untuk itulah dibutuhkan pendekatan yang mampu memisahkan dan mengorganisasi informasi berdasarkan bagian atau karakteristik tertentu dari produk. Dengan cara ini, setiap opini dapat ditautkan pada bagian produk yang spesifik, sehingga pemahaman terhadap kelebihan dan kekurangan suatu produk menjadi lebih terarah dan informatif[7].

Untuk mengatasi tantangan dalam memahami beragam opini yang terkandung dalam ulasan produk, dibutuhkan pendekatan analitis yang tidak hanya mampu mengungkap sentimen secara umum, tetapi juga dapat menelusuri opini tersebut hingga ke aspek-aspek spesifik yang dibahas. Salah satu pendekatan yang relevan untuk tujuan ini adalah *Aspect Based Sentiment Analysis* (ABSA). ABSA merupakan metode dalam analisis teks yang dirancang untuk mengelompokkan informasi berdasarkan aspek tertentu dari suatu objek seperti harga, kualitas, atau kemasan serta mengidentifikasi sentimen yang berkaitan dengan masing-masing aspek tersebut[8]. Dengan demikian, pendekatan ini memberikan gambaran yang lebih rinci dan terstruktur, sehingga sangat bermanfaat bagi konsumen dan produsen dalam memahami persepsi pengguna secara lebih mendalam terhadap bagian-bagian tertentu dari sebuah produk.

Pentingnya pendekatan ini juga tercermin dalam sejumlah penelitian terdahulu yang telah menerapkan ABSA untuk menganalisis ulasan konsumen dalam konteks yang berbeda. Salah satunya ditunjukkan oleh penelitian Gracia Radiena dan Adi

Nugroho (2023) yang berfokus pada ulasan pengguna terhadap aplikasi KAI Access. Penelitian ini menyoroti pentingnya mengelompokkan opini berdasarkan aspek-aspek seperti pemesanan tiket dan kualitas layanan untuk memahami kepuasan pengguna secara lebih spesifik[9]. Temuan tersebut menguatkan bahwa pendekatan ABSA dapat membantu mengidentifikasi aspek yang paling memengaruhi persepsi konsumen, sehingga relevan juga diterapkan pada ulasan produk kecantikan yang memiliki keragaman opini serupa[9].

Dalam penerapannya, proses ABSA meliputi dua langkah pokok, yaitu identifikasi aspek (*aspect extraction*) dan analisis sentimen terhadap aspek yang telah diketahui tersebut (*sentiment analysis*)[7]. Tahap pertama berfokus pada menemukan aspek-aspek spesifik yang dibahas dalam ulasan, sehingga analisis sentimen dapat dilakukan secara lebih terperinci pada masing-masing aspek [10]. Untuk mendukung proses ini, salah satu pendekatan yang dapat dimanfaatkan adalah *Topic modeling*, yaitu teknik yang mampu mengidentifikasi pola-pola tersembunyi dalam sekumpulan teks. Dengan memecah kumpulan data teks yang sangat besar menjadi kelompok-kelompok Topik yang lebih kecil dan terfokus, metode ini membantu mempermudah pemahaman dan analisis teks dalam skala besar secara lebih sistematis dan terstruktur[11].

Salah satu metode yang sering digunakan dalam penerapan *Topic modeling* adalah *Latent Dirichlet Allocation* (LDA). Metode ini bekerja dengan mengasumsikan bahwa setiap dokumen merupakan campuran dari beberapa Topik tersembunyi[12]. LDA memungkinkan pengelompokan kata-kata dalam ulasan menjadi tema-tema tertentu tanpa perlu informasi awal mengenai isi dokumen [12]. Dalam konteks ulasan produk kecantikan, LDA dapat membantu mengidentifikasi Topik seperti kualitas, harga, kemasan, dan lainnya, sehingga mempermudah proses ekstraksi aspek dalam ABSA secara lebih sistematis.

Setelah aspek dalam ulasan berhasil diidentifikasi, tahap berikutnya dalam proses ABSA adalah menentukan sentimen yang terkait dengan masing-masing aspek tersebut. Proses ini membutuhkan metode klasifikasi untuk mengelompokkan ulasan berdasarkan jenis sentimen yang terkandung. Oleh karena itu, diperlukan metode klasifikasi yang efektif dan akurat. Salah satu metode yang umum digunakan dalam analisis sentimen adalah *Logistic Regression*.

Logistic Regression merupakan algoritma pembelajaran mesin yang digunakan untuk memprediksi kemungkinan suatu data masuk ke dalam kategori tertentu. Dalam konteks analisis sentimen, kategori ini biasanya berupa sentimen positif dan negatif. *Logistic Regression* menggunakan fungsi logistik untuk mengubah data teks menjadi nilai probabilitas yang kemudian digunakan dalam pengambilan keputusan klasifikasi. Keunggulannya terletak pada proses yang sederhana, cepat, serta hasil prediksi yang mudah dipahami karena berbasis probabilitas.

Pada penelitian yang dilakukan oleh Hanif Reangga Alhakiem dan Erwin Budi Setiawan (2023)[13] dilakukan analisis sentimen berbasis aspek (*Aspect Based Sentiment Analysis*(ABSA) terhadap tweet pengguna terkait Telkomsel. Penelitian ini memfokuskan pada dua aspek utama, yaitu signal dan service. Metode klasifikasi yang digunakan adalah *Logistic Regression*, dengan dukungan ekspansi fitur menggunakan *FastText* untuk meminimalkan ketidaksesuaian kosakata dalam data teks. Sebanyak 16.987 tweet digunakan dalam penelitian ini, dengan tambahan data dari korpus IndoNews untuk memperkuat proses ekspansi fitur. Hasil pengujian menunjukkan bahwa pendekatan ini mampu meningkatkan performa klasifikasi. *F1-score* terbaik untuk aspek signal mencapai 96,48% dan untuk aspek service sebesar 95,57%, masing-masing diperoleh dari fitur-fitur teratas hasil ekspansi menggunakan *FastText*[13].

Selanjutnya, Penelitian yang dilakukan oleh Dinda Adimanggala et al. (2021)[14] menggunakan metode *Latent Dirichlet Allocation* (LDA) untuk melakukan ekstraksi aspek dari 28.222 ulasan pada platform *e-Commerce*. Penelitian ini mengevaluasi relevansi Topik tersembunyi yang dihasilkan LDA dengan membandingkannya terhadap penilaian manusia. Evaluasi dilakukan menggunakan *coherence score* dan pengujian *word vector*. Hasil menunjukkan bahwa LDA mampu menghasilkan Topik yang cukup relevan, dengan rata-rata selisih nilai koherensi sebesar 6%. Hal ini membuktikan bahwa meskipun identifikasi manusia masih lebih unggul, LDA tetap dapat membantu dalam menentukan Topik ketika koherensi aspek rendah[14].

Berdasarkan penelitian sebelumnya, analisis sentimen berbasis aspek pada studi ini mengombinasikan *Latent Dirichlet Allocation* (LDA) dan *Logistic Regression* dengan peran yang terpisah. LDA digunakan untuk mengekstraksi dan

memetakan aspek atau Topik yang muncul secara otomatis pada ulasan (*aspect extraction*). Keluaran LDA tidak dimanfaatkan sebagai fitur pada tahap klasifikasi. Selanjutnya, *Logistic Regression* diterapkan untuk menentukan polaritas sentimen pada teks yang telah dipetakan ke masing-masing aspek, sehingga menghasilkan analisis sentimen berbasis aspek yang lebih terstruktur dan dapat ditafsirkan. Kombinasi metode ini diharapkan dapat menghasilkan analisis sentimen berbasis aspek yang lebih akurat dan informatif, sehingga dapat membantu pihak terkait dalam memahami persepsi pengguna secara lebih mendalam terhadap berbagai aspek pada produk.

1.1.1 Permasalahan

1. Ketergantungan pada anotasi manusia membuat menghambat kelancaran proses pelabelan aspek dalam jumlah data yang besar. Tidak memiliki label aspek secara eksplisit sehingga menyulitkan identifikasi aspek secara manual.
2. Klasifikasi sentimen pada ulasan produk menghadapi tantangan karena data teks memiliki dimensi fitur yang tinggi dan hubungan antarkata yang kompleks.

1.2 Rumusan Masalah

1.2.1 Metode Usulan

1. Untuk mengidentifikasi aspek yang dibahas dalam ulasan yang tidak memiliki label eksplisit, digunakan metode *Latent Dirichlet Allocation* (LDA). Metode ini bekerja secara *unsupervised* dengan cara mengelompokkan kata berdasarkan distribusi kemunculannya dalam dokumen. Melalui proses tersebut, LDA mampu mengekstraksi Topik-Topik yang merepresentasikan aspek utama dalam ulasan tanpa memerlukan pelabelan manual.
2. Setelah aspek berhasil diidentifikasi, tahap selanjutnya adalah klasifikasi sentimen menggunakan metode *Logistic Regression*. Metode ini digunakan untuk menentukan polaritas sentiment pada setiap aspek yang telah diidentifikasi dari setiap ulasan. Dengan demikian, hasil klasifikasi memberikan gambaran apakah ulasan membahas aspek bersifat positif atau negative.

1.2.2 Pertanyaan Penelitian

Berdasarkan uraian permasalahan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Apa aspek-aspek utama yang dapat diidentifikasi menggunakan metode *Latent Dirichlet Allocation* pada ulasan konsumen?
2. Berapa tingkat akurasi yang dihasilkan oleh algoritma *Logistic Regression* dalam mengklasifikasikan polaritas sentimen pada ulasan untuk setiap subset aspek?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan fokus, terdapat beberapa batasan yang diterapkan dalam penelitian ini, yaitu sebagai berikut:

1. Data ulasan yang digunakan dibatasi hanya pada teks berbahasa Indonesia; ulasan dalam bahasa asing tidak disertakan dalam analisis.
2. Penelitian ini tidak mencakup kajian perbandingan antar metode dalam tahap ekstraksi aspek maupun klasifikasi sentimen.
3. Kategori sentimen yang digunakan dalam klasifikasi hanya mencakup dua kelas, yaitu positif dan negatif.
4. Analisis tidak mempertimbangkan jenis atau kategori produk kecantikan secara spesifik.
5. Penilaian pengguna dalam bentuk rating bintang tidak dijadikan variabel dalam proses analisis sentimen.
6. Data ulasan yang digunakan berjumlah 3.200 ulasan dengan rentang waktu tahun 2020 hingga 2025.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan

Penelitian ini memiliki tujuan yang ingin dicapai sebagai berikut:

1. Untuk mengidentifikasi aspek-aspek utama yang terkandung dalam ulasan pengguna *Female Daily* terkait produk kecantikan dengan menerapkan metode pemodelan Topik *Latent Dirichlet Allocation* (LDA), sehingga aspek yang paling sering dibahas dapat diketahui secara lebih terstruktur.
2. Untuk mengukur akurasi algoritma *Logistic Regression* dalam mengklasifikasikan sentimen pengguna, sehingga dapat diperoleh gambaran mengenai ketepatan model dalam menentukan kecenderungan opini atau sentimen terhadap aspek yang dibahas.

1.4.2 Manfaat

Adapun manfaat yang diharapkan dari penelitian ini antara lain:

1. Penelitian ini dapat memperluas wawasan dalam bidang *text mining* dan *machine learning*, khususnya terkait penerapan teknik *Aspect Based Sentiment Analysis* dalam domain ulasan konsumen.
2. Hasil dari penelitian ini dapat dimanfaatkan oleh pelaku industri kecantikan untuk mengidentifikasi aspek produk yang paling memengaruhi persepsi dan kepuasan pengguna, sehingga dapat mendukung perbaikan produk maupun strategi pemasaran yang lebih tepat sasaran.

1.5 Sistematika Penulisan

Penulisan laporan penelitian ini disusun dalam lima bab utama dengan penjelasan sebagai berikut:

1. BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang penelitian yang mendasari pentingnya analisis sentimen berbasis aspek pada ulasan produk kecantikan di *Female Daily*. Selain itu, bab ini juga berisi rumusan masalah, batasan masalah, tujuan penelitian, serta manfaat penelitian bagi berbagai pihak.

2. BAB II KAJIAN PUSTAKA

Bab ini menguraikan teori-teori yang mendukung penelitian, termasuk konsep analisis sentimen berbasis aspek, *preprocessing*, serta metode *Latent Dirichlet Allocation* (LDA) dan *Logistic Regression*. Selain itu, dibahas pula penelitian terdahulu yang relevan untuk memberikan landasan bagi metode yang digunakan dalam penelitian ini.

3. BAB III METODE PENELITIAN

Bab ini menjelaskan secara rinci metode yang digunakan dalam penelitian, mencakup sumber dan teknik pengumpulan data, proses *preprocessing text*, ekstraksi aspek menggunakan *Latent Dirichlet Allocation*, serta klasifikasi sentimen menggunakan *Logistic Regression*. Selain itu, dijelaskan pula teknik evaluasi model menggunakan *akurasi*, *precision*, *recall*, *F1-score*.

4. BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian yang meliputi *preprocessing* data, ekstraksi aspek menggunakan LDA, dan klasifikasi sentimen menggunakan *Logistic Regression*. Bab ini juga menjelaskan skenario pengujian, seperti pembagian data latih dan data uji, jumlah Topik LDA, serta konfigurasi *Logistic Regression* yang digunakan. Selain itu, ditampilkan hasil evaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*, beserta pembahasannya untuk melihat efektivitas metode yang diterapkan.

5. BAB V KESIMPULAN DAN SARAN

Bab ini berisi ringkasan dari keseluruhan hasil penelitian yang telah diperoleh, termasuk kesimpulan mengenai penerapan LDA dalam ekstraksi aspek dan *Logistic Regression* dalam klasifikasi sentimen pada ulasan *Female Daily*. Selain itu, bab ini juga memuat saran yang dapat dijadikan acuan untuk penelitian selanjutnya, baik dalam pengembangan metode, peningkatan kualitas data, maupun penerapan teknologi yang lebih mutakhir untuk analisis sentimen berbasis aspek.