

BAB I PENDAHULUAN

1.1.Latar Belakang

Di era digital saat ini, industri kosmetik mengalami perkembangan yang sangat pesat, tidak hanya secara global, tetapi juga di Indonesia. Kosmetik kini tidak hanya dianggap sebagai kebutuhan sekunder, melainkan telah menjadi bagian dari gaya hidup masyarakat, baik bagi perempuan maupun laki-laki. Kosmetik merupakan campuran bahan yang diaplikasikan ke bagian luar tubuh seperti kulit, rambut, dan bibir, yang bertujuan untuk memperindah, melindungi, dan memperbaiki penampilan seseorang[1]. Industri kecantikan di Indonesia mengalami pertumbuhan yang pesat setiap tahunnya. Hal ini tercermin dari peningkatan jumlah perusahaan kecantikan sebesar 21,9% dibandingkan tahun 2022, yang tercatat sebanyak 913 perusahaan, dan meningkat menjadi 1.010 perusahaan pada tahun 2023, atau meningkat 10 kali lipat dalam 5 tahun terakhir. Seiring dengan pertumbuhan industri kecantikan, masyarakat semakin aktif dalam memberikan ulasan atau review terhadap produk yang mereka gunakan melalui berbagai platform daring[2].

Ulasan terhadap produk kecantikan tersebut, kini menjadi salah satu acuan penting dalam proses pengambilan keputusan sebelum membeli, karena dapat memberikan gambaran mengenai kualitas, efektivitas, aroma, hingga kemasan produk tersebut. Ulasan produk kecantikan mampu merepresentasikan kepuasan atau kekecewaan pengguna, yang secara tidak langsung mencerminkan performa produk itu sendiri. Salah satu platform yang secara khusus menyediakan ruang bagi masyarakat untuk memberikan ulasan terhadap produk-produk kecantikan atau kosmetik adalah *Female Daily*[3].

Female Daily merupakan platform kecantikan terkemuka di Indonesia yang menyediakan artikel dan review produk kecantikan dari lebih dari 865.000 anggota yang tersebar di seluruh wilayah[1]. Platform *Female Daily* menjadi tempat bagi para konsumen di Indonesia dalam mencari informasi, membandingkan produk, dan mengetahui pengalaman pengguna lain secara lebih detail. Pada website ini, pengguna memiliki kebebasan untuk menyampaikan pengalaman pribadi mereka

selama menggunakan produk kosmetik melalui komentar teks. Meskipun tersedia sistem penilaian berbentuk skor bintang, tidak jarang ditemukan ketidaksesuaian antara jumlah bintang yang diberikan dan isi ulasan komentar[4]. Di dalam situs ini, setiap kategori produk memiliki fitur pengurutan berdasarkan berbagai kriteria, salah satunya adalah tingkat popularitas. Produk yang tergolong populer biasanya ditandai dengan jumlah ulasan yang tinggi. Namun demikian, tingginya jumlah ulasan tidak selalu mencerminkan kualitas positif dari produk tersebut. Dalam beberapa kasus, produk yang banyak diulas justru memiliki dominasi sentimen negatif, yang menunjukkan bahwa produk tersebut kurang direkomendasikan meskipun memiliki eksposur tinggi di platform[1]. Selain itu, banyaknya ulasan yang tersedia tidak selalu memudahkan dalam memperoleh informasi yang relevan, karena tidak semua ulasan terstruktur dengan baik dan konsisten. Masalah ini menimbulkan kebutuhan akan teknik analisis yang tidak hanya dapat mengklasifikasikan opini menjadi positif ataupun negatif, tetapi juga mampu mengidentifikasi aspek-aspek tertentu yang menjadi perhatian pengguna, seperti aroma, harga, tekstur, dan kemasan[5]. Oleh karena itu dibutuhkan teknik pendekatan dalam memahami ulasan pengguna secara lebih terperinci berdasarkan aspek-aspek tertentu dari produk yang dibahas.

Salah satu teknik pendekatan yang digunakan adalah analisis sentimen berbasis aspek (*Aspect-Based Sentiment Analysis/ABSA*)[6]. Pendekatan ini bertujuan untuk menganalisis setiap ulasan dengan mengidentifikasi berbagai aspek atau komponen produk, lalu menentukan sentimen yang sesuai positif atau negatif untuk masing-masing aspek tersebut[7]. Misalnya, dalam satu ulasan, pengguna dapat memberikan sentimen positif terhadap aroma produk namun memberikan sentimen negatif terhadap kemasannya. Dengan menggunakan pendekatan ini, informasi yang diperoleh menjadi lebih jelas dan spesifik karena dapat menunjukkan aspek mana yang paling banyak mendapat perhatian pengguna serta bagaimana persepsi mereka terhadap aspek tersebut[8]. Maka dari itu diperlukan teknik bantu yang mampu mengidentifikasi dan memahami topik atau aspek-aspek yang dominan dalam ulasan produk di platform *Female Daily*.

Topic Modeling merupakan salah satu teknik dalam *text mining* yang digunakan untuk menemukan dan mengelompokkan topik-topik yang belum diketahui dari sekumpulan data teks tanpa memerlukan label atau kategori sebelumnya[7]. Beberapa algoritma yang umum digunakan dalam topic modeling antara lain *Latent Dirichlet Allocation* (LDA), *Non-negative Matrix Factorization* (NMF), dan *Latent Semantic Analysis* (LSA)[9].

Pada penelitian ini menggunakan pendekatan *topic modeling* yaitu *Latent Dirichlet Allocation* karena mampu mengolah data teks yang berdimensi tinggi dengan cara menyederhanakan representasi kata ke dalam bentuk topik[10]. Metode ini bekerja dengan cara menghasilkan sejumlah topik yang memiliki bobot tertentu dalam setiap dokumen, sehingga sangat berguna untuk meringkas isi dokumen, menemukan pola topik, serta menghubungkan dokumen yang memiliki kesamaan isi[11]. LDA memungkinkan pengelompokan topik secara otomatis dari kumpulan teks ulasan, sehingga dapat mengidentifikasi aspek-aspek utama yang sering dibahas oleh pengguna dalam review produk di platform *Female Daily*[12]. Dengan menggunakan LDA, dapat diketahui topik dominan yang digunakan dalam penelitian ini, yakni: harga produk, tekstur, aroma, dan kemasan [3]. Pendekatan ini sangat membantu dalam mengenali fokus opini konsumen atau pengguna tanpa perlu memberikan label secara manual pada setiap aspek dalam ulasan.

Salah satu metode klasifikasi yang dapat digunakan dalam analisis sentimen berbasis aspek adalah *Long Short-Term Memory*. LSTM dikembangkan sebagai solusi untuk keterbatasan *Recurrent Neural Network*, terutama dalam memproses data dengan urutan panjang[13]. Metode ini sangat sesuai dan efektif untuk digunakan dalam klasifikasi teks, karena mampu memberikan performa yang baik pada berbagai bahasa, termasuk bahasa Indonesia. Dengan menggunakan sel memori internal, LSTM mampu menyimpan informasi dalam jangka waktu yang lebih lama, sehingga dapat memahami hubungan dependensi antar kata yang tersebar dalam banyak langkah waktu pada sebuah teks[14].

Arsitektur LSTM memiliki unit yang lebih kompleks dibandingkan dengan unit pada RNN standar. Kompleksitas ini memang membuat proses pelatihan model

membutuhkan waktu yang lebih lama. Namun, seiring dengan kekurangan tersebut, LSTM juga memiliki kemampuan *learning* yang lebih baik, khususnya dalam menangkap pola dan hubungan jangka panjang pada data teks[15]. Meskipun LSTM efektif dalam menangani dependensi jangka panjang dalam data sekuensial, metode ini masih dapat mengalami kendala ketika dihadapkan pada konteks yang sangat bervariasi atau berubah-ubah[16].

Penelitian yang dilakukan oleh [17] berjudul "*Enhancing Sentiment Analysis of Online Comments: A Novel Approach Integrating Topic Modeling and Deep Learning*" mengusulkan pendekatan gabungan antara *Latent Dirichlet Allocation* dan *Long Short-Term Memory* untuk meningkatkan akurasi analisis sentimen pada komentar daring dalam skala besar. Pendekatan ini dirancang untuk mengatasi keterbatasan model klasifikasi konvensional yang sering kali gagal menangkap konteks semantik yang kompleks dalam teks opini yang panjang dan tidak terstruktur. Penelitian ini menggunakan dua dataset besar dan menguji performa berbagai kombinasi metode. Hasil evaluasi menunjukkan bahwa model gabungan LDA dan LSTM memberikan tingkat akurasi yang sangat tinggi, yaitu mencapai 97,8% pada dataset SMILE dan 97,0% pada dataset Kaggle. Selain akurasi, model ini juga mencatat nilai precision, recall, dan F1-score yang tinggi, mengungguli metode baseline seperti Naive Bayes, Random Forest, dan SVM.

Berdasarkan penelitian oleh [18] untuk "*Sentiment Analysis of Indonesia's Capital Relocation Using Word2vec and Long Short-Term Memory Method*" menggunakan dua arsitektur *Word2Vec*, yaitu *Skip-gram* dan CBOW, serta beberapa tuning hyperparameter LSTM seperti fungsi aktivasi *ReLU*, *Sigmoid*, *Tanh* dan *optimizer RMSProp*, *Adam*. Hasil menunjukkan bahwa kombinasi *Skip-gram* dengan dimensi vektor 300, fungsi aktivasi *ReLU*, dan *optimizer RMSProp* mampu menghasilkan akurasi tertinggi sebesar 91%, dibandingkan konfigurasi lainnya. Selain akurasi yang baik, penelitian ini juga menyoroti adanya kendala seperti *overfitting*, yang muncul dari perbedaan kinerja antara data pelatihan dan validasi meskipun tetap efektif untuk menangani dataset yang kompleks.

Studi penelitian yang dilakukan oleh [7] membahas analisis sentimen berbasis aspek terhadap ulasan tempat wisata di Kalimantan Tengah menggunakan kombinasi metode *Latent Dirichlet Allocation* dan *Long Short-Term Memory*. LDA digunakan untuk mengekstraksi aspek-aspek utama dari ulasan, seperti lokasi, alam, pengalaman, dan fasilitas, sedangkan LSTM digunakan untuk mengklasifikasikan sentimen terhadap aspek tersebut. Hasil penelitian menunjukkan bahwa metode ini cukup efektif, dengan model LSTM mencapai akurasi sebesar 84%, presisi 84%, recall 100%, dan F1-score sebesar 91%. Studi ini menunjukkan bahwa kombinasi LDA dan LSTM mampu menangani data teks yang tidak terstruktur dan memberikan hasil klasifikasi sentimen yang baik berdasarkan aspek spesifik dalam ulasan pariwisata.

Penelitian terkait juga dari [19] menggunakan metode *Long Short-Term Memory* untuk melakukan klasifikasi sentimen berbasis aspek pada ulasan destinasi wisata Indonesia. LSTM dipilih karena kemampuannya dalam mengolah data teks berurutan dan mengenali dependensi jangka panjang dalam ulasan. Berdasarkan hasil evaluasi, model LSTM mencapai akurasi sebesar 90,63% dan F1-score sebesar 70,42% pada klasifikasi sentimen. Hasil ini menunjukkan bahwa LSTM cukup efektif dalam menangani ulasan berbahasa Indonesia yang bersifat tidak terstruktur.

1.2. Rumusan Masalah

1.2.1. Permasalahan

1. Ulasan pengguna terhadap produk kecantikan di platform *Female Daily* umumnya bersifat bebas, tidak terstruktur, dan sering kali membahas dengan tanpa adanya konteks aspek. Hal ini membuat proses identifikasi aspek secara manual dan aspek yang mendominasi menjadi kurang maksimal.
2. Panjang teks dan variasi kata ulasan pada platform *Female Daily* berpotensi menimbulkan masalah hilangnya informasi penting, yang dikenal sebagai *vanishing gradient*. Kondisi ini dapat mengurangi kemampuan klasifikasi untuk mempertahankan konteks dari awal hingga akhir ulasan, sehingga sulit

mengenali hubungan antara bagian ulasan yang membahas suatu aspek dengan bagian yang menyampaikan pendapat atau sentimen.

1.2.2. Metode Usulan

1. Menggunakan *Latent Dirichlet Allocation* (LDA) untuk mengekstraksi dan mengelompokkan topik atau aspek dominan dari ulasan pengguna yang bersifat bebas dan tidak terstruktur. Metode ini dapat membantu menemukan konteks aspek secara otomatis tanpa memerlukan pelabelan manual, sehingga mengatasi kesulitan dalam mengidentifikasi aspek yang sering dibahas.
2. Menerapkan *Long Short-Term Memory* (LSTM) untuk melakukan klasifikasi sentimen pada setiap aspek yang telah diketahui. LSTM dipilih karena kemampuannya mempertahankan konteks pada teks panjang dan mengurangi risiko *vanishing gradient*, sehingga hubungan antara bagian ulasan yang membahas aspek dan bagian yang menyampaikan sentimen dapat dianalisis lebih akurat.

1.2.3. Pertanyaan Penelitian

1. Bagaimana mengekstraksi topik-topik yang muncul dari opini masyarakat mengenai ulasan produk kecantikan di platform *Female Daily*?
2. Berapa tingkat akurasi dari parameter *Long Short-Term Memory* untuk mengklasifikasikan sentimen berdasarkan aspek-aspek yang telah diidentifikasi?

1.3. Batasan Masalah

Batasan-batasan masalah yang ada dalam penelitian ini adalah:

1. Penelitian ini berfokus pada ulasan teks pengguna yang diambil dari platform *Female Daily* dan tidak mencakup data dari platform lain dengan dataset yang dapat diakses melalui link berikut: bit.ly/dataset_femaledaily. Dataset pada penelitian ini berjumlah 3216 ulasan dalam rentang waktu tahun 2020 sampai 2025.
2. Penelitian ini mengolah data ulasan berbahasa Indonesia dan tidak mencakup ulasan bahasa lain untuk menjaga konsistensi proses dan hasil penelitian.

3. Metode ekstraksi fitur yang digunakan dalam penelitian ini dibatasi pada *Word2Vec*, dan tidak dilakukan perbandingan dengan metode ekstraksi fitur lain.
4. Metode ekstraksi aspek yang digunakan dalam penelitian ini dibatasi pada *Latent Dirichlet Allocation* dan *Long Short-Term Memory* untuk klasifikasi sentimen tanpa perbandingan metode lain.
5. Setiap ulasan mempertimbangkan aspek dalam level dokumen yang diidentifikasi, yakni: aspek tekstur, harga, aroma, efek, dan kemasan yang secara otomatis menggunakan teknik *aspect extraction*, tanpa mempertimbangkan keberadaan unit kalimat secara terpisah.
6. Klasifikasi sentimen dalam penelitian ini dibatasi pada dua kategori, yaitu sentimen positif dan negatif.

1.4. Tujuan dan Manfaat

1.4.1. Tujuan

Tujuan yang ingin didapatkan berdasarkan rumusan masalah pada penelitian ini adalah:

1. Mengidentifikasi dan mengelompokkan aspek-aspek tertentu dari ulasan produk *skincare* di platform *Female Daily* dengan menggunakan permodelan topik *Latent Dirichlet Allocation* (LDA).
2. Untuk mengklasifikasikan sentimen melalui pengukuran metrik evaluasi performa *accuracy*, *precision*, *recall*, dan *F1-score*, berdasarkan kombinasi metode *Latent Dirichlet Allocation* (LDA) dan *Long Short-Term Memory* (LSTM).

1.4.2. Manfaat

Adapun manfaat dalam penelitian ini adalah sebagai berikut:

1. Bagi penulis, penelitian ini merupakan sebuah eksplorasi teori-teori yang selama ini dipelajari, serta menambah wawasan, ilmu pengetahuan, dan pengalaman terhadap analisis sentimen berbasis aspek mengenai ulasan konsumen.

2. Bagi Universitas, sebagai tolak ukur pengetahuan mahasiswa dalam menguasai ilmu yang sudah dipelajari dan sebagai referensi untuk penelitian selanjutnya.
3. Bagi pembaca, memberikan informasi mengenai analisis sentimen berbasis aspek pada ulasan produk *Female Daily* dan bermanfaat untuk referensi penelitian di bidang Teknik Informatika atau sejenisnya.
4. Bagi pelaku industri kecantikan, hasil penelitian ini dapat memberikan gambaran mengenai opini dan sentimen pengguna terhadap produk kecantikan yang ditinjau berdasarkan aspek-aspek tertentu, sehingga dapat dijadikan masukan dalam meningkatkan kualitas produk dan strategi pemasaran yang tepat sasaran.

1.5. Sistematika Penulisan

Penyusunan penelitian tugas akhir ini disusun secara sistematis dalam 5 bab, dimulai dari BAB I sampai BAB V. Berikut penjelasan pada masing-masing bab:

1. BAB I PENDAHULUAN

Bab ini berisi tentang penjelasan latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, keaslian penelitian, dan sistematika penulisan penelitian.

2. BAB II KAJIAN PUSTAKA

Bab ini berisi penjelasan tentang tinjauan pustaka dan landasan teori yang berhubungan dengan penelitian ini.

3. BAB III METODE PENELITIAN

Bab ini berisi penjelasan tentang tahap-tahap atau metode yang digunakan untuk mencapai tujuan dan memperoleh hasil penelitian.

4. BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil dan pembahasan penelitian yang diperoleh dari seluruh tahapan metode penelitian yang telah dilakukan.

5. BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang disusun berdasarkan hasil dan pembahasan penelitian yang telah dilakukan secara keseluruhan.