

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Produk *skincare* merupakan salah satu kebutuhan penting dalam menjaga kesehatan dan penampilan kulit. Di Indonesia, pertumbuhan industri *skincare* sangat pesat seiring meningkatnya kesadaran masyarakat akan pentingnya perawatan diri. Salah satu merek populer adalah Hada Labo, yang dikenal dengan formulasi sederhana dan fokus pada kelembapan kulit. Produk *Hada Labo Gokujyun Ultimate Moisturizing Face Wash* menjadi salah satu varian unggulan karena diklaim mampu menjaga kelembapan kulit tanpa menimbulkan iritasi, sehingga banyak digunakan oleh konsumen dengan berbagai jenis kulit (Aini, *et. al.*, 2024:24–26).

Perkembangan era digital turut mengubah cara konsumen mencari informasi sebelum membeli produk *skincare*. Platform berbasis komunitas online seperti *Female Daily* kini menjadi sumber utama bagi konsumen untuk memperoleh ulasan langsung dari pengguna lain. Ulasan ini dianggap lebih jujur dan relevan dibandingkan promosi produsen, karena mencerminkan pengalaman nyata konsumen (Berlianti dan Hidayat, 2024). Oleh karena itu, ulasan di *Female Daily* berperan besar dalam membentuk persepsi dan keputusan pembelian produk kecantikan, termasuk Hada Labo.

Namun, jumlah ulasan yang sangat banyak menimbulkan masalah baru. Konsumen maupun produsen sulit membaca dan memahami semua ulasan secara manual, padahal ulasan tersebut memuat informasi berharga mengenai kualitas, kelebihan, maupun kekurangan produk. Kondisi ini menuntut adanya metode analisis otomatis yang dapat mengklasifikasikan ulasan ke dalam kategori sentimen positif, netral, atau negatif secara cepat dan akurat. Tanpa adanya sistem tersebut, wawasan dari konsumen berpotensi tidak dimanfaatkan secara optimal (Pramesti, 2024; Hariyanto, 2020:109).

Sebagai solusi, penelitian ini mengembangkan sistem analisis distribusi sentimen ulasan produk *skincare* dengan memanfaatkan kombinasi metode Naive

Bayes dan *Natural language Processing* (NLP). Proses dimulai dengan pelabelan otomatis menggunakan IndoBERT, model bahasa berbasis transformer yang efektif untuk teks bahasa Indonesia (Koto, *et. al.*, 2020). Selanjutnya, ulasan direpresentasikan dalam bentuk numerik menggunakan TF-IDF lalu diklasifikasikan dengan Naive Bayes berbasis NLP dan dilakukan evaluasi tingkat akurasi untuk mengklasifikasikan sentimen ulasan. Hasil analisis ditampilkan melalui aplikasi interaktif berbasis Streamlit lengkap dengan visualisasi distribusi sentimen dan WordCloud. Dengan pendekatan ini, penelitian diharapkan dapat memberikan kontribusi dalam memahami persepsi konsumen sekaligus membantu produsen dalam merumuskan strategi pemasaran produk *skincare*.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, penelitian ini bertujuan untuk menjawab beberapa pertanyaan berikut:

1. Bagaimana distribusi sentimen ulasan pengguna terhadap produk Hada Labo Gokujyun Ultimate Moisturizing Face Wash pada *platform Female Daily*?
2. Bagaimana pengaruh pelabelan otomatis menggunakan IndoBERT terhadap hasil klasifikasi Naive Bayes?
3. Seberapa tinggi tingkat akurasi algoritma Naive Bayes Berbasis *Natural Language Processing* (NLP) dalam mengklasifikasikan sentimen ulasan setelah menggunakan pelabelan otomatis?
4. Bagaimana merancang visualisasi interaktif hasil analisis sentimen menggunakan Streamlit?

1.3 Batasan Masalah

Penelitian ini akan dibatasi pada beberapa aspek untuk menjaga fokus dan kejelasan dalam pembahasan, yaitu:

1. Data ulasan yang digunakan hanya berasal dari *platform Female Daily* dan difokuskan pada produk Hada Labo Gokujyun Ultimate Moisturizing Face Wash.
2. Analisis sentimen diklasifikasikan ke dalam tiga kategori utama, yaitu: positif, negatif, dan netral.
3. Proses *preprocessing* data mencakup tahapan dengan NLP sebagai tools

pendukung berupa normalisasi, *case folding*, *tokenization*, *stopword removal*, dan *stemming* menggunakan pustaka Sastrawi.

4. Pelabelan data dilakukan secara otomatis menggunakan model IndoBERT, bukan secara manual.
5. Algoritma yang digunakan untuk klasifikasi adalah Naive Bayes dari pustaka *Scikit-learn*.
6. Sistem yang dikembangkan diimplementasikan dalam bentuk aplikasi web interaktif menggunakan framework Streamlit.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis distribusi sentimen ulasan pengguna terhadap produk *Hada Labo Gokujyun Ultimate Moisturizing Face Wash* pada platform *Female Daily*
2. Menganalisis pengaruh pelabelan otomatis menggunakan IndoBERT terhadap hasil klasifikasi Naive Bayes.
3. Mengevaluasi tingkat akurasi algoritma Naive Bayes berbasis NLP dalam mengklasifikasikan sentimen ulasan setelah menggunakan pelabelan otomatis.
4. Mengembangkan aplikasi dan mengimplementasikan visualisasi interaktif hasil analisis sentimen menggunakan Streamlit.

1.5 Kegunaan Tugas Akhir

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Kegunaan akademik

- a. Menambah referensi dalam bidang *text mining* dan analisis sentimen, khususnya pada penerapan algoritma Naive Bayes dengan pelabelan otomatis menggunakan model IndoBERT.
- b. Memberikan kontribusi terhadap pengembangan metode klasifikasi teks berbahasa Indonesia dengan memanfaatkan kombinasi *machine learning* dan *deep learning*.
- c. Menjadi bahan rujukan untuk penelitian sejenis di bidang analisis sentimen dan pemrosesan bahasa alami (*Natural language Processing*).

2. Kegunaan praktis

- a. Memberikan solusi otomatis bagi konsumen dan pelaku industri *skincare* untuk memahami persepsi pasar terhadap suatu produk melalui analisis ulasan pengguna.
- b. Menyediakan aplikasi berbasis web yang interaktif, memudahkan pengguna dalam mengunggah data, melakukan analisis, dan melihat visualisasi hasil secara real-time.
- c. Membantu produsen dalam merancang strategi pemasaran yang lebih efektif berdasarkan analisis sentimen yang akurat.

1.6 Tempat Tugas Akhir

Penelitian ini dilaksanakan dengan menggunakan data yang diperoleh secara daring tanpa keterlibatan langsung pihak tersebut. *Platform Female Daily*, yaitu sebuah situs komunitas dan *review* produk kecantikan yang populer di Indonesia. Ulasan yang dikumpulkan berasal dari pengguna yang memberikan pendapat mereka secara langsung mengenai produk Hada Labo Gokujyun Ultimate Moisturizing Face Wash. Seluruh proses pengumpulan dan pengolahan data dilakukan secara mandiri oleh peneliti tanpa keterlibatan langsung dari pihak *Female Daily*.

1.7 Waktu Pelaksanaan Tugas Akhir

Penelitian ini dilaksanakan mulai bulan April hingga Juli 2025. Pada bulan April dilakukan penyusunan proposal dan konsultasi awal. Pengumpulan serta pra- pemrosesan data dilakukan selama bulan Mei hingga Juni.

1.8 Metodologi Tugas Akhir

Langkah-langkah metodologi yang digunakan dalam penelitian ini meliputi:

1. Pengumpulan data ulasan produk dari situs *Female Daily* menggunakan teknik *web scraping*.
2. *Preprocessing* data teks yang mencakup tahapan normalisasi, *case folding*, *tokenization*, *stopword removal*, dan *stemming* menggunakan pustaka Sastrawi.
3. Pelabelan otomatis sentimen menggunakan model IndoBERT dengan tiga kategori: positif, negatif, dan netral.

4. Ekstraksi fitur dari teks berlabel menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*).
5. Pelatihan dan pengujian model klasifikasi menggunakan algoritma Naive Bayes dari pustaka *Scikit-learn*.
6. Evaluasi performa model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*.
7. Implementasi hasil analisis ke dalam aplikasi berbasis Streamlit, dengan fitur-fitur sebagai berikut:
 - a. *Upload file* CSV berisi ulasan pengguna.
 - b. *Preprocessing* otomatis terhadap teks ulasan.
 - c. Pelabelan sentimen otomatis menggunakan model IndoBERT.
 - d. Pelatihan dan evaluasi model Naive Bayes.
 - e. Visualisasi hasil berupa *WordCloud* dan grafik distribusi sentimen.
 - f. Unduh hasil klasifikasi ke dalam file CSV.

1.9 Sistematika Penulisan

Penulisan tugas akhir ini disusun secara sistematis untuk memudahkan pembaca dalam memahami alur pemikiran, metodologi, dan hasil penelitian yang dilakukan. Adapun sistematika penulisan dalam tugas akhir ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, lokasi dan waktu penelitian, metodologi, serta sistematika penulisan tugas akhir.

BAB II TINJAUAN PUSTAKA

Bab ini membahas teori yang mendukung penelitian, meliputi konsep analisis sentimen, algoritma Naive Bayes, Natural Language Processing (NLP), model IndoBERT, serta teknologi pendukung. Selain itu, disajikan ringkasan penelitian terdahulu yang relevan dengan topik.

BAB III ANALISIS DAN PERANCANGAN

Bab ini menguraikan analisis sistem, meliputi tujuan sistem, prosedur yang berjalan, kebutuhan data, analisis fungsional, analisis pengguna, serta evaluasi menggunakan SWOT. Analisis ini menjadi dasar perancangan sistem yang sesuai.

BAB IV IMPLEMENTASI SISTEM

Bab ini menyajikan perancangan dan implementasi sistem berbasis Python menggunakan Streamlit. Perancangan mencakup prosedur, data, arsitektur, dan antarmuka dengan diagram pendukung. Bab ini juga menampilkan hasil implementasi, uji coba aplikasi, serta evaluasi performa model dengan metrik akurasi, precision, recall, dan F1-score.

BAB V PENUTUP

Bab terakhir ini berisi kesimpulan dari hasil penelitian serta implementasi sistem yang telah dilakukan. Selain itu, disampaikan pula saran-saran yang dapat digunakan sebagai pertimbangan untuk penelitian atau pengembangan sistem lebih lanjut di masa mendatang.

