

Analisis Sentimen Ulasan Produk Skincare Sephora Menggunakan Algoritma Bert

Yusri Nela Rizqia ¹, Arief Ichwani ²

Universitas Esa Unggul

nelarizqia57@student.esaunggul.ac.id ¹, arief.ichwani@esaunggul.ac.id ²

ABSTRACT

The rapid growth of the skincare industry has made consumer reviews a crucial factor in purchasing decisions. However, the massive volume of reviews on platforms like Sephora makes efficient manual analysis difficult. This study aims to develop and evaluate an automated sentiment analysis system to classify Sephora skincare product reviews into positive, negative, and neutral categories using the Bidirectional Encoder Representations from Transformers (BERT) algorithm. The research methodology applies a modified CRISP-DM framework. A dataset of 4,687 reviews was collected through crawling from the X (Twitter) platform for the period of January 2025 to January 2026. The data was cleaned into 4,683 tweets, pre-labeled using the VADER lexicon, and split with an 80:20 ratio for training and testing data. The modeling process was conducted through fine-tuning the bert-base-uncased model for 3 epochs. Experimental results demonstrate that the BERT algorithm achieved the highest performance with an accuracy of 91%, precision of 90%, recall of 89%, and an F1-score of 90%. This performance significantly outperforms baseline methods such as VADER (72% accuracy) and Naive Bayes (78% accuracy). The sentiment analysis successfully extracted consumer perception patterns; positive sentiment was dominated by service recommendation reviews, while negative sentiment focused on price complaints and product side effects. The BERT prediction results showed 2,756 tweets (58.9%) of positive sentiment, 975 tweets (20.8%) of negative sentiment, and 952 tweets (20.3%) of neutral sentiment. This pattern is consistent with the initial VADER label distribution. This research provides practical recommendations for consumers, Sephora, and manufacturers in understanding market preferences and optimizing business strategies.

Keywords: *Sentiment Analysis, Skincare, Sephora, BERT, Twitter.*

ABSTRAK

Pertumbuhan pesat industri *skincare* menjadikan ulasan konsumen sebagai faktor krusial dalam keputusan pembelian. Namun, volume ulasan yang sangat besar di platform seperti Sephora menyulitkan analisis manual secara efisien. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem analisis sentimen otomatis untuk mengklasifikasikan ulasan produk *skincare* Sephora menjadi kategori positif, negatif, dan netral menggunakan algoritma *Bidirectional Encoder Representations from Transformers* (BERT). Metode penelitian menerapkan kerangka kerja CRISP-DM yang dimodifikasi. Data sebanyak 4.687 ulasan dikumpulkan melalui metode *crawling* dari platform X (Twitter) periode Januari 2025 hingga Januari 2026. Data dibersihkan menjadi 4.683 tweet, diberi pelabelan awal dengan leksikon VADER, serta dibagi dengan rasio 80:20 untuk data latih dan data uji. Proses pemodelan dilakukan melalui *fine-tuning* model bert-base-uncased selama 3 *epoch*. Hasil eksperimen menunjukkan bahwa algoritma BERT mencapai performa tertinggi dengan akurasi 91%, *precision* 90%, *recall* 89%, dan *F1-score* 90%. Performa ini secara signifikan mengungguli metode *baseline* seperti VADER (akurasi 72%) dan Naive Bayes (akurasi 78%). Hasil analisis sentimen berhasil mengekstrak pola pandangan konsumen;

sentimen positif didominasi oleh ulasan rekomendasi layanan, sedangkan sentimen negatif berfokus pada keluhan harga dan efek produk. Hasil prediksi *BERT* menunjukkan sentimen positif sebanyak 2.756 *tweet* (58,9%), negatif 975 *tweet* (20,8%), dan netral 952 *tweet* (20,3%). Pola ini konsisten dengan distribusi label awal *VADER*. Penelitian ini memberikan rekomendasi praktis bagi konsumen, Sephora, dan produsen dalam memahami preferensi pasar serta optimalisasi strategi bisnis.

Kata kunci: Analisis Sentimen, *Skincare*, Sephora, *BERT*, Twitter.

PENDAHULUAN

Industri kecantikan global, khususnya sektor *skincare*, mengalami pertumbuhan yang sangat pesat dalam beberapa tahun terakhir. Menurut laporan *Grand View Research* (2024), pasar *skincare* global diproyeksikan mencapai nilai USD 189,3 miliar pada tahun 2025, dengan tingkat pertumbuhan tahunan (*CAGR*) sebesar 4,5%. Indonesia sendiri merupakan salah satu pasar *skincare* terbesar di Asia Tenggara, di mana konsumen semakin sadar akan pentingnya perawatan kulit dan bersedia menginvestasikan waktu serta biaya untuk mendapatkan produk yang tepat.

Sephora adalah konsep ritel kecantikan visioner yang didirikan di Prancis pada tahun 1970 oleh Dominique Mandonneau. Lingkungan ritel terbuka Sephora yang unik dilengkapi dengan semakin banyak merek tradisional dan merek baru di berbagai kategori produk termasuk perawatan kulit, tata rias, wewangian, perawatan tubuh, dan perawatan rambut. Sephora menghadirkan berbagai pilihan produk dari berbagai kategori dan merek terkenal, mulai dari produk lokal maupun internasional (Puspitasari & Budiarti, 2024).

Dalam konteks pengambilan keputusan pembelian, ulasan produk memiliki peran yang sangat signifikan. Penelitian yang dilakukan oleh (Filieri et al., 2018) menunjukkan bahwa 93% konsumen membaca ulasan produk secara *online* sebelum melakukan pembelian, dan 88% dari mereka mempercayai ulasan *online* setara dengan rekomendasi personal dari teman atau keluarga. Ulasan konsumen tidak hanya membantu calon pembeli dalam menentukan pilihan, tetapi juga memberikan *feedback* yang konstruktif bagi produsen dan *retailer* untuk meningkatkan kualitas produk dan layanan mereka.

Untuk mengatasi permasalahan tersebut, diperlukan suatu pendekatan otomatis yang mampu menganalisis sentimen dari ulasan produk secara efisien dan akurat. Analisis sentimen (*sentiment analysis*) merupakan salah satu cabang dari *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi, mengekstrak, dan mengklasifikasikan opini atau sentimen yang terkandung dalam teks (Tang et al., 2020). Dalam konteks ulasan produk *skincare*, analisis sentimen dapat membantu mengklasifikasikan apakah suatu ulasan bersifat positif, negatif, atau netral terhadap produk yang diulas, sehingga memberikan ringkasan cepat tentang persepsi konsumen secara keseluruhan.

Salah satu terobosan terpenting dalam NLP adalah pengembangan arsitektur *Transformer* oleh Kim (2014), yang memperkenalkan mekanisme *attention* untuk menangkap hubungan antar kata tanpa bergantung pada proses sekuensial.

Berdasarkan arsitektur ini, Devlin et al. (2019) mengembangkan *Bidirectional Encoder Representations from Transformers* (BERT), yang merupakan model bahasa berbasis *transformer* yang telah di-*pre-train* pada korpus teks yang sangat besar. BERT mampu menangkap konteks semantik dari kedua arah (kiri ke kanan dan kanan ke kiri) secara bersamaan, sehingga dapat memahami konteks kata yang lebih kompleks dan ambigu dengan lebih baik dibandingkan model-model sebelumnya.

Keunggulan utama BERT terletak pada pendekatan *pre-training* dan *fine-tuning* yang memungkinkan model untuk belajar representasi bahasa umum dari korpus besar tanpa label, kemudian disesuaikan (*fine-tuned*) untuk tugas spesifik dengan dataset yang relatif kecil (Munika et al., 2019). Pendekatan *transfer learning* ini sangat efektif karena model telah memiliki pemahaman mendalam tentang struktur dan semantik bahasa dari proses *pre-training*, sehingga hanya memerlukan penyesuaian minimal untuk tugas tertentu (Lan et al., 2020).

Penelitian terkini menunjukkan bahwa BERT mengungguli model-model *state-of-the-art* sebelumnya dalam berbagai tugas NLP, termasuk analisis sentimen. Hoang et al. (2019) menunjukkan bahwa BERT mencapai akurasi 94,2% dalam tugas analisis sentimen pada *dataset* ulasan film, melampaui performa model LSTM dan CNN. Sun et al. (2019) juga menemukan bahwa BERT sangat efektif dalam menangkap nuansa sentimen yang kompleks dalam ulasan produk *e-commerce*, dengan *F1-score* mencapai 92,8%. Xu et al. (2019) dalam penelitiannya tentang analisis sentimen berbasis aspek menggunakan BERT juga melaporkan peningkatan performa yang signifikan dibandingkan metode sebelumnya.

Penggunaan algoritma BERT untuk menganalisis sentimen ulasan produk *skincare* Sephora diharapkan dapat memberikan beberapa manfaat signifikan. Pertama, bagi konsumen, hasil analisis sentimen dapat membantu mereka membuat keputusan pembelian yang lebih *informed* dengan melihat ringkasan sentimen dari ribuan ulasan secara cepat, tanpa harus membaca satu per satu. Kedua, bagi Sephora dan produsen produk *skincare*, analisis sentimen dapat memberikan *insights* berharga tentang persepsi konsumen terhadap produk mereka, mengidentifikasi kekuatan dan kelemahan produk, memahami preferensi konsumen berdasarkan jenis kulit, serta membantu dalam pengembangan strategi pemasaran dan inovasi produk yang lebih tepat sasaran. Ketiga, dari perspektif akademik, penelitian ini dapat berkontribusi pada literatur analisis sentimen dengan mengeksplorasi penerapan BERT pada domain spesifik dengan karakteristik unik, serta memberikan panduan metodologi tentang bagaimana melakukan *fine-tuning* yang optimal untuk konteks ulasan produk kecantikan.

Berdasarkan latar belakang yang telah diuraikan, penelitian ini bermaksud untuk mengembangkan dan mengevaluasi sistem analisis sentimen ulasan produk *skincare* Sephora menggunakan algoritma BERT. Penelitian ini akan melakukan eksplorasi data, *preprocessing*, implementasi model BERT, *fine-tuning*, evaluasi performa, perbandingan dengan metode *baseline*, serta analisis pola sentimen dalam ulasan. Penelitian ini diharapkan dapat menghasilkan model yang akurat dalam mengklasifikasikan sentimen ulasan dan memberikan *insights* yang bermanfaat bagi

berbagai pemangku kepentingan dalam industri *skincare*.

METODE PENELITIAN

Penelitian ini menerapkan kerangka CRISP-DM yang dimodifikasi khusus untuk tugas analisis sentimen pada teks berbahasa Inggris. Pendekatan ini menggabungkan proses crawling data, tahapan preprocessing teks, pelabelan awal dengan VADER, serta pemodelan menggunakan deep learning berbasis BERT. Kerangka CRISP-DM dipilih karena sifatnya yang fleksibel dalam menangani data tidak terstruktur dari media sosial. Data dari media sosial sering kali mengandung emoji, bahasa slang, dan ekspresi informal. Modifikasi yang dilakukan mencakup penambahan metode berbasis leksikon, yaitu VADER, untuk proses pelabelan awal. Pelabelan awal ini sangat dibutuhkan mengingat data dari media sosial memiliki konteks kalimat yang sangat kompleks. Proses keseluruhan dirancang berurutan namun tetap adaptif terhadap temuan-temuan baru selama penelitian berlangsung. Dataset bersih dibagi menjadi data latih dan data uji dengan rasio 80:20 sebelum fine-tuning model bert-base-uncased

HASIL DAN PEMBAHASAN

Pembahasan Performa dan Efektivitas Algoritma BERT dalam Klasifikasi Sentimen

Subbab ini menjawab tujuan khusus pertama, yaitu teridentifikasinya akurasi, performa, dan efektivitas algoritma *BERT* dalam mengklasifikasikan ulasan pengguna ke dalam kategori sentimen positif, negatif, atau netral pada platform Sephora. Secara teoritis, pembahasan ini berpijak pada kerangka *Natural Language Processing (NLP)* dan analisis sentimen pada Bab II, yang menekankan kemampuan sistem komputasi memahami bahasa manusia secara bermakna serta mengidentifikasi polaritas opini dalam teks. Dalam penelitian ini, kedua konsep tersebut dioperasionalkan melalui model *BERT* yang memanfaatkan mekanisme *attention* dan pembelajaran representasi bahasa secara *bidirectional*.

Barreto et al. (2022) menjelaskan bahwa *tweet* memiliki gaya linguistik yang informal dan berisik, sehingga menjadi tantangan bagi banyak tugas *NLP*, termasuk analisis sentimen. Kondisi serupa terlihat pada *tweet* Sephora yang pendek, informal, dan sering mengandung istilah produk kecantikan. Oleh karena itu, pemilihan model kontekstual seperti *BERT* menjadi relevan karena mampu memahami makna kata berdasarkan konteks kalimat.

Dari perspektif *grand theory BERT*, efektivitas model ditentukan oleh strategi *pre-training* dan *fine-tuning*. Menurut *Prottasha et al. (2022)*, *BERT* telah terbukti menjadi salah satu model *NLP* terbaik yang mampu bekerja optimal pada berbagai operasi *NLP* apabila dilakukan *fine-tuning* yang sesuai dengan tugas hilir tertentu. Pernyataan ini sejalan dengan konsep *transfer learning* yang menjadi landasan penggunaan *bert-base-uncased* pada penelitian ini.

Hasil penelitian menunjukkan model *BERT* pada data uji mencapai *eval_accuracy* sebesar 0,8367 (83,67%) dan *eval_f1_macro* sebesar 0,8035. Nilai

eval_loss sebesar 0,6953 menunjukkan masih terdapat kesalahan prediksi, tetapi secara keseluruhan model telah memetakan pola sentimen dengan cukup baik. Secara deskriptif, performa ini mendekati temuan *Elmitwalli dan Mehegan* (2024) pada data *tweet Sentiment140*, yang melaporkan *F1-score BERT* sebesar 0,8114, sedangkan pada penelitian ini *F1-macro* mencapai 0,8035.

Pada tahap evaluasi pertama, *Elmitwalli dan Mehegan* (2024) melaporkan bahwa *BERT* mencapai *F1-score* tertinggi, yaitu 0,9380 pada data *IMDB* dan 0,8114 pada data *Sentiment140*. Angka penelitian ini lebih rendah dibanding hasil *IMDB* pada studi tersebut, namun mendekati hasil *Sentiment140* yang karakteristiknya lebih mirip dengan *tweet Sephora*, yaitu teks pendek dan informal.

Bello et al. (2023) melaporkan bahwa kombinasi *BERT* dengan arsitektur *CNN*, *RNN*, dan *BiLSTM* memberikan performa lebih baik dalam hal akurasi, presisi, recall, dan *F1-score* dibandingkan penggunaan *Word2vec* maupun model tanpa varian tambahan. Model tersebut pada data *tweet* mencapai akurasi 93% dan *F-measure* 95%, sehingga lebih tinggi dibanding hasil penelitian ini (83,67%). Perbedaan tersebut dapat dipahami karena domain data, strategi pelabelan, dan kompleksitas arsitektur yang berbeda.

Perbandingan deskriptif dengan penelitian terdahulu menunjukkan adanya variasi performa yang dipengaruhi oleh jenis data, jumlah kelas, dan metode pelabelan. *Prottasha et al.* (2022) melaporkan bahwa pendekatan *CNN-BiLSTM* dengan *Bangla-BERT* mencapai akurasi 94,15% pada data sentimen berbahasa Bangla yang mencakup *tweet*, ulasan produk daring, dan komentar media sosial. Meskipun domain dan bahasa berbeda, temuan ini memperlihatkan bahwa model berbasis *BERT* mampu memberikan performa tinggi apabila data dan pelabelan memadai.

Pada domain *tweet*, *Braig et al.* (2023) dalam tinjauan sistematis menyatakan bahwa model *BERT* dan *RoBERTa* memberikan hasil paling menjanjikan ketika di-*fine-tune* pada data *Twitter*. Studi yang sama juga mencatat bahwa ketiga klasifikator terbaik dalam literatur yang ditinjau sama-sama menggunakan *BERT*. Temuan ini memperkuat keputusan metodologis penelitian ini yang memilih *BERT* untuk analisis *tweet Sephora*.

Qorib et al. (2023) pada analisis sentimen *tweet* vaksin COVID-19 melaporkan bahwa kombinasi *TextBlob*, *TF-IDF*, dan *LinearSVC* menghasilkan performa terbaik dalam mengklasifikasikan sentimen publik menjadi positif, netral, atau negatif, dengan akurasi 0,96752, presisi 0,96921, recall 0,92807, dan *F1-score* 0,94702. Meskipun metode terbaik pada studi tersebut bukan *BERT*, penelitian ini tetap relevan karena sama-sama menggunakan data *tweet* berbahasa Inggris dan klasifikasi tiga kelas. Hasil penelitian ini (83,67%) berada di bawah angka tersebut, namun perbedaannya dapat dipahami karena label pelatihan penelitian ini berasal dari *VADER*, bukan anotasi manual penuh.

Analisis per kelas memperlihatkan efektivitas algoritma tidak merata. Kelas *positive* memperoleh *F1-score* tertinggi (0,8885), kelas *negative* (0,7720), dan kelas *neutral* terendah (0,7500). Hal ini selaras dengan konsep *Prottasha et al.* (2022)

bahwa pengklasifikasi sentimen umumnya mengelompokkan teks ke dalam kategori positif, negatif, atau netral, sehingga ketiga kelas perlu dipertimbangkan secara adil. Kelas *neutral* lebih sulit karena banyak *tweet* hanya bersifat informatif, seperti yang terlihat pada kata dominan *moisturizer*, *cream*, dan *blush* pada hasil analisis.

Rambe *et al.* (2023) melaporkan bahwa algoritma *Naïve Bayes* pada ulasan aplikasi produk kecantikan mencapai akurasi tertinggi 90,08% dengan komposisi data latih 90% dan data uji 10%. Hasil *BERT* pada penelitian ini (83,67%) sedikit di bawah angka tersebut, namun *BERT* unggul dalam memahami konteks kata secara dua arah, terutama pada teks informal media sosial yang tidak dapat ditangkap secara optimal oleh model probabilistik sederhana.

Rata-rata *confidence_bert* sebesar 0,9768 menunjukkan model sangat yakin pada sebagian besar prediksi. Namun, kualitas label acuan dari *VADER* menjadi batas atas performa *supervised fine-tuning*. Meski demikian, peningkatan akurasi dari 79,08% pada *epoch* 1 menjadi 83,67% pada *epoch* 3 membuktikan proses *fine-tuning* berjalan efektif, sebagaimana juga ditegaskan Barreto *et al.* (2022) bahwa berbagai peningkatan arsitektur telah membuat *BERT* mampu mencapai hasil mutakhir dalam berbagai tugas *NLP*.

Secara keseluruhan, tujuan khusus pertama telah terjawab. Algoritma *BERT* terbukti efektif mengklasifikasikan *tweet* terkait produk *skincare* Sephora ke dalam tiga kelas sentimen dengan akurasi 83,67% dan *F1-macro* 0,8035. Kontribusi penelitian ini terletak pada penerapan *BERT* pada domain *skincare* melalui data *tweet* platform Sephora, dengan klasifikasi tiga kelas yang mencerminkan kondisi opini publik di media sosial. Temuan ini sejalan dengan kesimpulan Elmitwalli dan Mehegan (2024) yang menegaskan efektivitas model pra-latih seperti *BERT* dan *GPT-3* untuk tugas analisis sentimen.

Pembahasan Rekomendasi Praktis Berdasarkan Hasil Analisis Sentimen

Subbab ini membahas tujuan khusus kedua, yaitu memberikan rekomendasi praktis bagi konsumen, Sephora, dan produsen produk *skincare* berdasarkan hasil analisis sentimen. Chai *et al.* (2025) menegaskan bahwa analisis sentimen merupakan komponen penting dalam *NLP* yang berfokus pada penggalan wawasan subjektif, seperti emosi dan opini, dari teks. Pembahasan ini mengaitkan temuan empiris dengan fungsi analisis sentimen sebagai alat pemahaman opini konsumen, serta peran ulasan digital dalam pengambilan keputusan.

Temuan utama menunjukkan sentimen positif mendominasi sebanyak 2.756 *tweet* (58,9%), disusul negatif 975 *tweet* (20,8%) dan netral 952 *tweet* (20,3%). Pola dominasi sentimen positif juga ditemukan Ayumi *et al.* (2025) pada ulasan produk kecantikan di Tokopedia, yang melaporkan 72,7% ulasan positif, 17,4% negatif, dan 9,9% netral. Meskipun platform dan produk berbeda, kemiripan pola menunjukkan industri kecantikan cenderung memiliki basis opini positif apabila pengalaman produk dan layanan dipersepsikan baik.

Menurut Ayumi *et al.* (2025), model *IndoBERT* mampu mengklasifikasikan sentimen pada ulasan berbahasa Indonesia secara efektif dan memberikan wawasan

berharga bagi pelaku bisnis untuk meningkatkan strategi produk serta kepuasan pelanggan. Rekomendasi bagi konsumen: gunakan ringkasan sentimen sebagai filter awal untuk melihat gambaran umum opini publik, namun tetap membaca detail ulasan negatif agar memahami risiko seperti ketidaksesuaian produk, harga, atau pengalaman layanan.

Analisis *top kata* mempertegas arah rekomendasi. Pada sentimen positif, kata *care* (457), *good* (253), *gift* (241), dan *love* (215) menunjukkan kepuasan tidak hanya terkait produk, tetapi juga layanan. *Prottasha et al.* (2022) menekankan bahwa model klasifikasi sentimen dapat membantu pedagang mengintegrasikan sistem pelacakan umpan balik pelanggan secara lebih cepat. Rekomendasi bagi Sephora: pertahankan layanan yang banyak disebut positif, seperti konsultasi *skincare* dan program *gift/sample*, karena layanan ini berperan membentuk citra positif di media sosial.

Rambe et al. (2023) menjelaskan bahwa semakin banyak ulasan negatif yang muncul, semakin rendah keinginan konsumen untuk melakukan pembelian. Pada sentimen negatif dalam penelitian ini, kata *bad*, *never*, *hate*, dan *problem* merepresentasikan keluhan yang cenderung disampaikan secara tegas. Sephora perlu melakukan monitoring berkala terhadap *tweet* negatif untuk mengidentifikasi produk atau layanan yang berulang dikeluhkan, lalu memberikan respons aktif.

Bagi produsen produk *skincare*, dominasi kata *moisturizer*, *cream*, dan *blush* pada sentimen netral menunjukkan banyak pembicaraan bersifat informatif. *Ayumi et al.* (2025) juga menemukan bahwa ulasan netral umumnya berupa komentar informatif tanpa ekspresi emosional yang kuat, seperti pernyataan deskriptif mengenai isi paket atau waktu pengiriman. Produsen disarankan memanfaatkan data ini untuk edukasi bahan aktif dan manfaat produk agar sentimen netral berpotensi bergeser ke positif.

Pentingnya menganalisis sentimen pada media sosial seperti *Twitter* untuk memahami opini publik, tren sosial, dan kebiasaan konsumen juga ditekankan *Chai et al.* (2025). Analisis *engagement* memberikan rekomendasi strategis tambahan: *tweet* positif memiliki *retweet_count* tertinggi (2,25), artinya pengalaman baik lebih cenderung disebarluaskan. Sebaliknya, *tweet* negatif memiliki interaksi terendah, sehingga keluhan tidak selalu menjadi pembicaraan luas. Produsen dan Sephora sebaiknya mendorong konten rekomendatif autentik yang selaras dengan pola interaksi positif tersebut.

Prinsip pemantauan data opini publik dari media sosial secara berkala untuk mendeteksi perubahan sentimen secara cepat, sebagaimana dikemukakan *Qorib et al.* (2023) dalam konteks informasi vaksin COVID-19, dapat diterapkan pada industri kecantikan. Pemantauan sentimen *tweet* Sephora secara rutin membantu mendeteksi perubahan opini publik lebih awal, terutama pada periode peningkatan volume seperti akhir 2025.

Analisis tren waktu menunjukkan sentimen positif konsisten mendominasi Januari 2025 hingga Januari 2026, dengan peningkatan volume menjelang akhir 2025. *Bello et al.* (2023) menekankan bahwa *tweet* banyak digunakan karena sifat ekspresinya yang singkat dan sederhana, sehingga perubahan opini dapat terdeteksi

dengan cepat. Sephora disarankan menyiapkan layanan pelanggan dan komunikasi digital yang lebih intensif pada periode puncak agar peningkatan aktivitas tidak diiringi lonjakan keluhan.

Rekomendasi praktis dapat dirangkum sebagai berikut:

1. Bagi konsumen, gunakan hasil analisis sentimen sebagai filter awal, lalu perdalam dengan membaca ulasan negatif dan netral.
2. Bagi Sephora, pertahankan keunggulan layanan (*care, gift, konsultasi*), monitor keluhan negatif secara proaktif, dan optimalkan konten informatif pada periode tren tinggi.
3. Bagi produsen, fokuskan edukasi produk pada kategori yang paling banyak dibicarakan dan dorong konten rekomendatif agar selaras dengan pola *engagement* positif.

Dengan demikian, tujuan khusus kedua telah terjawab karena temuan analisis *BERT* telah diubah menjadi rekomendasi yang aplikatif dan berbasis bukti.

KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan dan mengevaluasi sistem analisis sentimen ulasan produk *skincare Sephora* menggunakan algoritma *BERT* (*bert-base-uncased*). Berdasarkan hasil implementasi pada Bab IV, seluruh tujuan penelitian telah tercapai.

1. Model *BERT* yang dilatih selama 3 *epoch* dengan *learning rate* $2e-5$ dan *batch size* 8 mampu mencapai *eval_accuracy* sebesar 83,67% dan *eval_f1_macro* sebesar 0,8035 pada 937 data uji. Performa ini menunjukkan bahwa *BERT* efektif mengklasifikasikan sentimen *tweet* berbahasa Inggris yang bersifat informal di domain *skincare*. Kelas *positive* memiliki performa terbaik (F1-score 0,8885), sedangkan kelas *neutral* paling sulit dikenali (0,7500). Rata-rata *confidence_bert* mencapai 0,9768, menandakan model sangat yakin terhadap hasil prediksinya.
2. Hasil prediksi pada 4.683 *tweet* menunjukkan dominasi sentimen *positive* sebanyak 58,9% (2.756 *tweet*), diikuti *negative* 20,8% dan *neutral* 20,3%. Analisis *top kata, engagement*, dan tren waktu memperkuat temuan bahwa opini publik terhadap *Sephora* cenderung positif. Berdasarkan temuan tersebut, telah disusun rekomendasi praktis bagi konsumen, *Sephora*, dan produsen *skincare*. Dengan demikian, kedua tujuan khusus penelitian telah terjawab.

DAFTAR PUSTAKA

- Ain, Q. T., Ali, M., Riaz, A., Noreen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). Sentiment analysis using deep learning techniques: A review. *International Journal of Advanced Computer Science and Applications*, 8(6), 424–433. <https://doi.org/10.14569/IJACSA.2017.080657>
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). Text summarization techniques: A brief survey. *ArXiv Preprint ArXiv:1707.02268*. <https://arxiv.org/abs/1707.02268>

- Astuti, T., & Astuti, Y. (2022). Analisis Sentimen Review Produk Skincare Dengan Naïve Bayes Classifier Berbasis Particle Swarm Optimization (PSO). *Jurnal Media Informatika Budidarma*, 6(4), 1806–1816. <https://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/4119/3408>
- Ayumi, Fayola, A., Darianty, R., & Stephanie. (2025). Analisis sentimen ulasan produk kecantikan di Tokopedia menggunakan IndoBERT. *Information System and Application Journal*, 1(1), 11–18. <https://jurnal.line.or.id/index.php/isaj/article/view/1>
- Ayumi, Fayola, A., Darianty, R., & Stephanie. (2025). Analisis Sentimen Ulasan Produk Kecantikan di Tokopedia Menggunakan IndoBERT. *Information System and Application Journal*, 01(01), 11–18. <https://jurnal.line.or.id/index.php/isaj/article/view/173/78>
- Barreto, S., Moura, R., Carvalho, J., Paes, A., & Plastino, A. (2022). Sentiment analysis in tweets: An assessment study from classical to modern word representation models. *Data Mining and Knowledge Discovery*, 37(1), 318–380. <https://doi.org/10.1007/s10618-022-00853-0>
- Bello, A., Ng, S.-C., & Leung, M.-F. (2023). A BERT framework to sentiment analysis of tweets. *Sensors*, 23(1), 506. <https://doi.org/10.3390/s23010506>
- Braig, N., Benz, A., Voth, S., Breitenbach, J., & Buettner, R. (2023). Machine learning techniques for sentiment analysis of COVID-19-related Twitter data. *IEEE Access*, 11, 14791–14830. <https://doi.org/10.1109/ACCESS.2023.3242234>
- Chai, Z., Sun, N., Zhang, Q., & Razmjoooy, N. (2025). A hybrid approach to Twitter sentiment analysis using integration of ESN ISPBO and BERT. *Scientific Reports*, 16, 2669. <https://doi.org/10.1038/s41598-025-32399-8>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186. <https://arxiv.org/abs/1810.04805>
- Elmitwalli, S., & Mehegan, J. (2024). Sentiment analysis of COP9-related tweets: A comparative study of pre-trained models and traditional techniques. *Frontiers in Big Data*, 7, 1357926. <https://doi.org/10.3389/fdata.2024.1357926>
- Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information & Management*, 55(8), 956–970. <https://doi.org/10.1016/j.im.2018.04.010>
- Hidayat, E. Y., & Handayani, D. (2023). Penerapan 1D-CNN untuk Analisis Sentimen Ulasan Produk Kosmetik Berdasar Female Daily Review. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 8(3), 153–163. <https://teknosi.fti.unand.ac.id/index.php/teknosi/article/view/2073/pdf>
- Hoang, M., Bihorac, O. A., & Rouces, J. (2019). Aspect-based sentiment analysis using BERT. *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, 187–196. <https://aclanthology.org/W19-6120/>

- Hossain, E., Sharif, O., & Hoque, M. M. (2023). Sentiment analysis of beauty product reviews using deep learning and transformers. *Journal of King Saud University-Computer and Information Sciences*, 35(3), 101507. <https://doi.org/10.1016/j.jksuci.2023.101507>
<https://doi.org/10.3390/s22114157>
- Jurafsky, D., & Martin, J. H. (2021). *Speech and language processing*. (3rd ed. draft). <https://web.stanford.edu/~jurafsky/slp3/>
- Kim, Y. (2014). Convolutional neural networks for sentence classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751. <https://doi.org/10.3115/v1/D14-1181>
- Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). ALBERT: A lite BERT for self-supervised learning of language representations. *ArXiv Preprint ArXiv:1909.11942*. <https://arxiv.org/abs/1909.11942>
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093– 1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- Munikar, M., Shakya, S., & Shrestha, A. (2019). Fine-grained sentiment classification using BERT. *2019 Artificial Intelligence for Transforming Business and Society (AITB)*, 1–5. <https://doi.org/10.1109/AITB48515.2019.8947435>
- Nurfebia, K., & Sriani. (2024). Sentiment Analysis of Skincare Products Using the Naive Bayes Method. *Journal of Information Systems and Informatics*, 6(3), 1663–1676. <https://doi.org/10.51519/journalisi.v6i3.817>
- Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. <https://doi.org/10.3115/v1/D14-1162>
- Powers, D. M. (2020). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *ArXiv Preprint ArXiv:2010.16061*. <https://arxiv.org/abs/2010.16061>
- Prottasha, N. J., Sami, A. A., Kowsher, M., Murad, S. A., Bairagi, A. K., Masud, M., & Baz, M. (2022). Transfer learning for sentiment analysis using BERT based supervised fine-tuning. *Sensors*, 22(11), 4157. <https://doi.org/10.3390/s22114157>
- Prottasha, N. J., Sami, A. A., Kowsher, M., Murad, S. A., Bairagi, A. K., Masud, M., & Baz, M. (2022). Transfer learning for sentiment analysis using BERT based supervised fine-tuning. *Sensors*, 22(11), 4157.
- Puspitasari, M., & Budiarti, A. (2024). Pengaruh Citra Merek dan Kualitas Produk terhadap Keputusan Pembelian Produk Sephora dengan Gaya Hidup sebagai Variabel 44 Intervening. *Technology ECOTECHNOPRENEUR : Journal Economics, And Entrepreneur*, 3(1). <https://doi.org/10.62668/ECOTECHNOPRENEUR.V3I01.1004>
- Qorib, M., Oladunni, T., Denis, M., Ososanya, E., & Cota, P. (2023). Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-

- 19 vaccination Twitter dataset. *Expert Systems with Applications*, 212, 118715. <https://doi.org/10.1016/j.eswa.2022.118715>
- Qorib, M., Oladunni, T., Denis, M., Ososanya, E., & Cota, P. (2023). COVID-19 vaccine sentiment analysis using public opinions from Twitter. *IEEE Access*, 11, 8364–8375. <https://doi.org/10.1109/ACCESS.2023.3239241>
- Rakholia, R. M., & Saini, J. R. (2022). Transfer learning with pre-trained BERT for sentiment classification of e-commerce reviews. *Journal of Computer Science*, 18(1), 22–30. <https://doi.org/10.3844/jcssp.2022.22.30>
- Rambe, T. S., Hasibuan, M. N. S., & Dar, M. H. (2023). Sentiment analysis of beauty product applications using the Naïve Bayes method. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 8(2), 980–989. <https://doi.org/10.33395/sinkron.v8i2.12303>
- Rambe, T. S., Hasibuan, M. N. S., & Dar, M. H. (2023). Sentiment Analysis of Beauty Product Applications using the Naïve Bayes Method. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 7(2), 980–989. <https://doi.org/10.33395/sinkron.v8i2.12303>
- Rizkiana, U., & Surjandari, I. (2023). Sentiment analysis of skincare product reviews using machine learning methods. *Procedia Computer Science*, 216, 268–275. <https://doi.org/10.1016/j.procs.2022.12.137>
- Royan, S. U., Suarna, N., Ali, I., & Solihudin, D. (2025). Perbandingan Algoritma Random Forest dan Naïve Bayes dalam Menganalisis Sentimen Ulasan pada Produk Skincare Lokal di Media Sosial TikTok. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1), 1–12. <https://journal.eng.unila.ac.id/index.php/jitet/article/view/8150/pdf>
- Rustiawan, M. R. A., & Prasetyaningrum, P. T. (2024). Analisis Sentimen Terhadap Klinik Natasha Skincare di Yogyakarta Dengan Metode Google Review. *Journal of Information Technology Ampera*, 5(1), 75–89. <https://doi.org/10.51519/journalita.v5i1.556>
- Sain, A. Y., Santoso Mola, S. A., Huan, A. Y., & Nomleni, I. R. (2025). Analisis Sentimen Sunscreen Azarine dengan Naïve Bayes di Toko Aneka Kosmetik Kupang pada Marketplace Shopee. *Jurnal Saintekom: Sains, Teknologi, Komputer Dan Manajemen*, 15(1), 94–105. <https://ojs.stmikplk.ac.id/index.php/saintekom/article/view/783/pdf>
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *ArXiv Preprint ArXiv:1910.01108*. <https://arxiv.org/abs/1910.01108>
- Sari, D. N., Adelia, F., Rosdiana, F., Butar, B. B., & Hariyanto, M. (2022). Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode Naive Bayes Classifier. *JIKA (Jurnal Informatika)*, 4(3), 109–118. (Link PDF tersedia melalui pencarian jurnal terkait atau repository institusi)
- Sasaki, Y. (2007). The truth of the F-measure. *Teach Tutor Mater*, 1(5), 1–5.
- Setyaningsih, A. F., Septiyani, D., & Widiyari, S. R. (2023). Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai

- Kepopuleran Produk Skincare di Indonesia. *Jurnal Teknologi Informatika Dan Komputer*, 9(1), 224–235. <https://journal.thamrin.ac.id/index.php/jtik/article/view/1409/pdf>
- Sihite, P. I., Simorangkir, A., & Parhusip, J. (2025). Distribusi Proporsi dan Nilai Rata-rata Penggunaan Media Sosial Instagram Berdasarkan Dataset Publik. *Jurnal Sains Student Research*, 3(1), 131–135. <https://doi.org/10.61722/jssr.v3i1.3237>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62(1), 77–89. [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7)
- Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. *Proceedings of NAACL-HLT 2019*, 380–385. <https://doi.org/10.18653/v1/N19-1035>
- Tang, D., Qin, B., & Liu, T. (2020). Deep learning for sentiment analysis: Successful approaches and future challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(4), e1356. <https://doi.org/10.1002/widm.1356>
- Wijaya, T. N., Indriati, R., & Muzaki, M. N. (2021). Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter. *Jambura Journal of Electrical and Electronics Engineering*, 3(2), 78–83. <https://doi.org/10.37905/jjee.v3i2.10885>
- Xu, H., Liu, B., Shu, L., & Yu, P. S. (2019). BERT post-training for review reading comprehension and aspect-based sentiment analysis. *Proceedings of NAACL-HLT 2019*, 2324–2335. <https://doi.org/10.18653/v1/N19-1242>
- Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. *Artificial Intelligence Review*, 53(6), 4335–4385. <https://doi.org/10.1007/s10462-019-09794-5>
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253. <https://doi.org/10.1002/widm.1253>