UNIVERSITAS
GADJAH MADA
INTEGRASI PEMODELAN TOPIK LDA, ANALISIS SENTIMEN BERT, DAN ALGORITMA PELUANG DALAM EKSTRAKSI SUARA PELANGGAN STUDI KASUS: PENGEMBANGAN PRODUK LIP CREAM
Vannencia Magdalena, Prof. Ir. Subagyo, Ph.D., IPU., ASEAN Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Industri kosmetik di Indonesia mengalami pertumbuhan pesat yang memicu persaingan ketat, khususnya pada kategori produk *lip cream.* Dalam lingkungan yang dinamis ini, perusahaan dituntut untuk memahami kebutuhan konsumen secara mendalam guna meminimalkan risiko kegagalan produk baru. Namun, ulasan konsumen pada *platform* publik seperti *Female Daily* sering kali hadir dalam jumlah masif dan format teks tidak terstruktur yang mengandung bahasa gaul (*slang*) serta singkatan, sehingga sulit dianalisis secara manual. Penelitian ini bertujuan untuk mengekstraksi "Suara Pelanggan" (*Voice of Customer*) secara otomatis dan akurat untuk mengidentifikasi atribut produk yang paling krusial bagi pengembangan dan strategi pemasaran *lip cream*.

Metode penelitian ini menggunakan pendekatan *big data analytics* dengan mengekstrak 47.737 ulasan konsumen dari 216 SKU pada *platform Female Daily Network* secara eksklusif dalam rentang tahun 2022 hingga 2025. Data dibersihkan melalui tahap *preprocessing* dan diuji kecukupan datanya menggunakan rumus tingkat ketelitian 5% dan kepercayaan 95%. Data kemudian dikelompokkan ke dalam empat kuadran klaster berdasarkan variabel rata-rata harga terstandardisasi per mililiter dan *rating* produk sebagai *industry baseline*. Proses ekstraksi VoC dilakukan menggunakan algoritma probabilistik *Latent Dirichlet Allocation* (LDA), yang kemudian diintegrasikan dengan model *Deep Learning* berbasis Transformer BERT pada level kalimat untuk menilai tingkat kepuasan pelanggan.

Hasil eksperimen komputasi menunjukkan performansi model LDA yang optimal dengan rata-rata *Coherence Score* pada rentang sangat baik (0,42 hingga 0,58) dan nilai tumpang tindih *Jaccard Similarity* yang rendah (< 0,13), menghasilkan 7 topik pada Klaster I, 7 topik pada Klaster II, 6 topik pada Klaster III, dan 8 topik pada Klaster IV. Model analisis sentimen BERT menunjukkan performansi klasifikasi yang tinggi dengan tingkat akurasi mencapai 67% dalam mendeteksi polaritas ulasan campuran. Evaluasi menggunakan Algoritma Peluang Ulwick mengungkap bahwa atribut Variasi Estetika (spektrum warna) secara konsisten berada di zona *underserved* lintas klaster dengan skor peluang tertinggi mencapai 17,471 pada segmen premium dan 12,130 pada segmen ekonomis, sementara aspek Pengalaman Aplikasi (desain fisik aplikator) menjadi fokus kritis *underserved* khusus pada Klaster III dengan skor kepuasan terendah sebesar 1,000.

Kata Kunci: *Pemodelan Topik, LDA, Analisis Sentimen, BERT, Opportunity Algorithm, Lip Cream*

UNIVERSITAS
GADJAH MADA
INTEGRASI PEMODELAN TOPIK LDA, ANALISIS SENTIMEN BERT, DAN ALGORITMA PELUANG DALAM EKSTRAKSI SUARA PELANGGAN STUDI KASUS: PENGEMBANGAN PRODUK LIP CREAM
Vannencia Magdalena, Prof. Ir. Subagyo, Ph.D., IPU., ASEAN Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

The cosmetics industry in Indonesia is experiencing rapid growth, triggering intense competition, particularly within the lip cream product category. In this dynamic environment, companies are compelled to deeply understand consumer needs to minimize the risk of new product failure. However, consumer reviews on public platforms such as Female Daily are often massive in volume and exist in unstructured text formats containing slang and abbreviations, making manual analysis highly difficult. This study aims to automatically and accurately extract the "Voice of Customer" (VoC) to identify the most crucial product attributes for lip cream development and marketing strategies.

The methodology of this research employs a big data analytics approach by exclusively extracting 47,737 consumer reviews from 216 Stock Keeping Units (SKUs) on the Female Daily Network platform within the 2022 to 2025 period. The data were cleaned through preprocessing stages and subjected to a data adequacy test utilizing a formula with a 5% precision level and a 95% confidence level. The data were then grouped into four cluster quadrants based on the variables of standardized average price per milliliter and product rating as the industry baseline. The VoC extraction process was conducted using the probabilistic Latent Dirichlet Allocation (LDA) algorithm, which was subsequently integrated with the Transformer-based Deep Learning model, BERT, at the sentence level to evaluate customer satisfaction rates.

The computational experiment results demonstrate optimal LDA model performance, with an average Coherence Score within an excellent range (0.42 to 0.58) and a low Jaccard Similarity overlap value (< 0.13), yielding 7 topics for Cluster I, 7 topics for Cluster II, 6 topics for Cluster III, and 8 topics for Cluster IV. The BERT sentiment analysis model exhibits high classification performance, achieving an accuracy rate of 67% in detecting the polarity of mixed reviews. Evaluation using Ulwick's Opportunity Algorithm reveals that the Aesthetic Variety (color spectrum) attribute consistently resides in the underserved zone across all clusters, with the highest opportunity scores reaching 17.471 in the premium segment and 12.130 in the economic segment. Meanwhile, the Application Experience (physical applicator design) aspect emerges as a critical underserved focus specifically within Cluster III, registering the lowest satisfaction score of 1.000.

Keywords: *Topic Modeling, LDA, Sentiment Analysis, BERT, Opportunity Algorithm, Lip Cream*