

ABSTRAK

Industri kecantikan Indonesia diproyeksikan mencapai nilai pasar USD 9,7 miliar pada 2025, namun persaingan yang ketat menyebabkan banyak brand lokal mengalami penurunan performa, termasuk Scarlett yang turun dari peringkat kedua pada 2022 ke luar dari sepuluh besar brand terpopuler pada 2024. Penelitian ini bertujuan mengidentifikasi atribut kualitas produk kecantikan berbasis persepsi konsumen dari ulasan *e-commerce* dan menyusun rekomendasi atribut untuk informasi produk di *platform* digital. Hasil tiga metode *text mining* diterapkan yaitu TF-IDF, *Latent Dirichlet Allocation* (LDA), dan BERTopic pada 135.000 ulasan konsumen dari 450 produk kecantikan terpopuler di Shopee, Tokopedia, dan Lazada. LDA ditetapkan sebagai metode utama dengan *coherence score* 0,571 pada K=6 karena menghasilkan topik yang lebih interpretatif dibandingkan BERTopic dengan 161 topik dan *coherence score* 0,696 namun kurang terstruktur. Hasil pemodelan mengidentifikasi enam topik utama yang dipetakan ke dalam empat dimensi *perceived quality* meliputi *intrinsic cues*, *extrinsic cues*, *experience attributes*, dan *credence attributes*. Berdasarkan pemetaan tersebut, produk *skincare* disarankan mengedepankan atribut lembab, cerah, dan jenis kulit, produk *makeup* mengedepankan *shade*, *coverage*, dan *waterproof*, serta produk perawatan tubuh mengedepankan wangi, aroma, dan tekstur dalam informasi produk di *platform e-commerce*.

Kata Kunci: *Text mining, E-commerce, Atribut Kualitas, LDA, Perceived Quality*

ABSTRACT

The Indonesian beauty industry is projected to reach a market value of USD 9.7 billion in 2025, yet intense competition has led many local brands to experience performance declines, including Scarlett which dropped from second place in 2022 to outside the top ten most popular brands in 2024. This study aims to identify beauty product quality attributes based on consumer perceptions derived from *e-commerce* reviews and formulate attribute recommendations for product information on digital platforms. Result three text mining methods, namely TF-IDF, *Latent Dirichlet Allocation* (LDA), and BERTopic, on 135,000 consumer reviews from 450 top beauty products on Shopee, Tokopedia, and Lazada. LDA was selected as the primary method with a coherence score of 0.571 at K=6 as it produced more interpretable topics compared to BERTopic which generated 161 topics with a coherence score of 0.696 but was less structured. The modeling results identified six main topics mapped into four perceived quality dimensions comprising *intrinsic cues*, *extrinsic cues*, *experience attributes*, and *credence attributes*. Based on this mapping, *skincare* products are recommended to emphasize moisturizing, brightening, and skin type attributes, *makeup* products to emphasize *shade*, *coverage*, and *waterproof* attributes, and body care products to emphasize scent, aroma, and texture attributes in product information on *e-commerce* platforms.

Keywords: *Text mining, E-commerce, Quality Attributes, LDA, Perceived Quality*