

Topic Modeling of Skincare Comments from Female Daily

Nabila¹, Chanifah Indah Ratnasari^{2*}

Program Studi Informatika, Universitas Islam Indonesia

nabila02@students.uii.ac.id¹, chanifah.indah@uii.ac.id²

* corresponding author

Article Info

Article history:

Received 2025-05-06
Revised 2025-07-01
Accepted 2025-07-03

Keyword:

Consumer review,
Female Daily,
LDA,
Skincare,
Topic modeling.

ABSTRACT

The increasing popularity of skincare products in Indonesia has encouraged many consumers to seek and share information through online platforms. One of the most influential platforms is Female Daily, which provides a space for users to review and discuss various skincare products. This study aims to explore the dominant topics within user-generated comments related to skincare products on Female Daily. The research employed a descriptive qualitative approach using topic modeling with Latent Dirichlet Allocation (LDA). Data were collected from user comments on several popular skincare products and were preprocessed through punctuation removal, case folding, tokenization, normalization, stopword removal, and stemming. The optimal number of topics was determined using coherence scores. The results reveal that users frequently discuss personal experiences, highlight product benefits and drawbacks, and often refer to their specific skin concerns. These insights provide valuable information for skincare brands to understand customer preferences and perceptions. In conclusion, topic modeling with LDA proves effective in extracting meaningful themes from large-scale textual data, offering a useful method for analyzing consumer feedback in the beauty industry.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Industri kosmetik merupakan salah satu penggerak utama perekonomian nasional [1]. Pertumbuhan industri ini di Indonesia sangat pesat mencapai 21,9%, yakni terdapat 913 perusahaan pada tahun 2022 dan pada pertengahan tahun 2023 terdapat 1.010 perusahaan [2]. Besarnya peluang bisnis kosmetik di Indonesia diimbangi oleh meningkatnya ancaman persaingan yang dihadapi oleh perusahaan kosmetik lokal [3]. Industri ini mencakup beragam produk kecantikan, termasuk produk perawatan kulit atau *skincare*. Dengan nilai pasar mencapai USD 2,05 miliar pada tahun 2022, produk *skincare* menjadi salah satu segmen yang tumbuh paling pesat dalam industri perawatan diri di Indonesia. Persaingan antar perusahaan juga semakin ketat seiring kemunculan merek-merek baru yang terus berinovasi untuk menarik konsumen [4].

Di era digital saat ini, keputusan pembelian konsumen, terutama untuk produk kosmetik, sangat dipengaruhi oleh ulasan *online*. Baik ulasan positif maupun negatif memiliki peran penting dalam membentuk persepsi konsumen terhadap suatu produk [5]. Temuan [6] juga mendukung hal ini,

menunjukkan bahwa konsumen cenderung mengandalkan pengalaman pribadi yang dibagikan di media sosial. Salah satu platform besar yang menjadi sumber ulasan dan diskusi terkait produk kecantikan adalah Female Daily. Platform ini merupakan forum daring yang memungkinkan pengguna untuk berbagi informasi, pengalaman, dan rekomendasi mengenai produk-produk kecantikan serta perawatan kulit [7].

Female Daily memiliki laman *website* dengan domain femaledaily.com dan versi aplikasi *mobile* bernama "Female Daily - Beauty Review" dengan jumlah 1 juta unduhan dan jumlah *member* pengguna yang terdaftar lebih dari 990 ribu. Female Daily menyediakan lebih dari 2,5 juta ulasan produk dari para anggotanya. Hingga saat ini, tercatat lebih dari 100 ribu produk telah direview, dengan kategori *skincare* sebagai yang paling dominan, yaitu lebih dari 1,35 juta ulasan [8]. Hal ini menunjukkan bahwa Female Daily merupakan platform review kecantikan terbesar di Indonesia.

Female Daily menyediakan berbagai fitur, seperti "FD Review" untuk ulasan produk, "FD Editorial" untuk konten artikel, "FD Studio" sebagai *e-commerce*, dan "FD Talk" sebagai forum diskusi. Pada "FD Review", pengguna dapat

mencari produk dan membaca informasi tentang ulasan mengenai komentar, tipe kulit, jangka pemakaian, dan penilaian bintang (rating) dari para pengguna mengenai sebuah produk yang telah mereka gunakan. Female Daily juga rutin menyelenggarakan survei setiap tahunnya yang hasilnya disajikan dalam ajang "Female Daily Best of Beauty Awards" (BoBA). Membaca semua ulasan yang tersedia akan memakan banyak waktu dan sering kali topik yang sedang dibahas tidak menggambarkan informasi secara keseluruhan.

Penelitian terdahulu mengenai *Latent Dirichlet Allocation* (LDA) telah dilakukan oleh [9] yang melakukan pemodelan topik pada media berita *online* untuk merek Somethinc. Penelitian tersebut memanfaatkan nilai *coherence* untuk menganalisis dan mengevaluasi jumlah topik terbaik yang terdapat pada topik 6 dengan nilai *coherence* sebesar 0,404. Hasil tersebut ditentukan berdasarkan nilai *coherence* tertinggi dengan melakukan *running* sebanyak 50 *passes* serta memperhatikan pendekatan penilaian manusia yang terbukti berpengaruh dalam menginterpretasikan topik. Dapat disimpulkan bahwa jumlah topik 6 menghasilkan topik-topik yang relevan mengenai merek Somethinc.

Penelitian [10] mengidentifikasi topik-topik dominan pada ulasan hotel yang memiliki sentimen positif maupun negatif menggunakan LDA dan PLSA (*Probabilistic Latent Semantic Analysis*), serta mengintegrasikan klasifikasi sentimen dengan *Naïve Bayes Classifier* (NBC). Studi ini menyimpulkan bahwa LDA memberikan kinerja yang lebih baik dibandingkan PLSA dalam mengidentifikasi topik laten pada data ulasan hotel, dibuktikan oleh nilai indeks validasi internal Calinski-Harabasz yang lebih tinggi pada model LDA.

Penelitian lain yang relevan dilakukan oleh [11] membandingkan metode klasifikasi dokumen teks ilmiah tidak terstruktur (*e-books*) berdasarkan teks lengkapnya dengan menerapkan pendekatan pemodelan topik LDA dan LSA (*Latent Semantic Analysis*). *Dataset* pada penelitian tersebut terdiri dari 100 buku yang berisi sekitar 1 juta kata berdasarkan teks lengkap. Hasil penelitian menunjukkan bahwa LDA memiliki hasil yang lebih baik dibandingkan LSA. Nilai *coherence* terbaik dengan jumlah topik sebanyak 20 yang diperoleh dari metode LDA adalah 0,54846, sedangkan nilai *coherence* LSA adalah 0,4047.

Sementara itu, beberapa studi telah mengkaji platform Female Daily dari berbagai sudut pandang. Penelitian oleh [12] meneliti pengaruh ulasan produk terhadap kepuasan pengguna aplikasi Female Daily. Penelitian ini menggunakan sampel *non probability sampling* dengan teknik *purposive sampling*. Penyeleksian sampel untuk penelitian ini mendapatkan 30 sampel dengan kriteria yang telah ditentukan. Hasil penelitian menunjukkan bahwa ulasan produk memiliki pengaruh yang kuat dan signifikan terhadap kepuasan pengguna, dengan nilai koefisien determinasi sebesar 72,9%.

Penelitian oleh [13] menganalisis sentimen terhadap ulasan produk *skincare* dari situs Female Daily dengan menerapkan metode *Support Vector Machine* (SVM). Data diperoleh

melalui proses *scraping* sebanyak 1.249 ulasan yang berasal dari produk-produk kecantikan lokal di Indonesia, dengan atribut utama berupa nama akun dan nama produk. Hasil analisis menunjukkan kinerja model yang cukup baik, dengan akurasi sebesar 87%, *recall* 90%, *precision* 84,90%, dan *f1-score* sebesar 87,37%. Namun demikian, sebagian besar studi sebelumnya yang menganalisis data dari Female Daily masih terbatas pada analisis sentimen menggunakan pendekatan *supervised* seperti SVM [13], atau fokus pada aspek rating dan *review* [12]. Penerapan metode *unsupervised* seperti *Latent Dirichlet Allocation* (LDA) dalam menggali struktur tematik komentar pengguna di platform ini, khususnya pada produk pemenang BoBA, masih jarang dilakukan. Hal ini menunjukkan adanya ruang untuk eksplorasi yang lebih mendalam mengenai preferensi dan persepsi konsumen melalui pendekatan *topic modeling*.

Berdasarkan pemaparan sebelumnya, terlihat bahwa metode LDA telah banyak diterapkan dan menunjukkan performa yang baik dalam mengidentifikasi topik utama dari data teks tidak terstruktur, baik dalam domain berita maupun *e-book*. LDA juga dapat mengatasi *overfitting* yang dialami PLSA [15]. Namun demikian, hingga saat ini belum ditemukan studi yang secara khusus menerapkan LDA untuk *topic modeling* pada ulasan produk di platform Female Daily. Hal ini menunjukkan adanya celah penelitian, mengingat Female Daily merupakan salah satu forum daring terbesar di Indonesia yang menjadi wadah utama opini konsumen terhadap produk kecantikan. Sebagian besar studi yang berfokus pada Female Daily masih terbatas pada analisis sentimen atau klasifikasi kepribadian, dan belum menjangkau eksplorasi topik laten dalam ulasan pengguna.

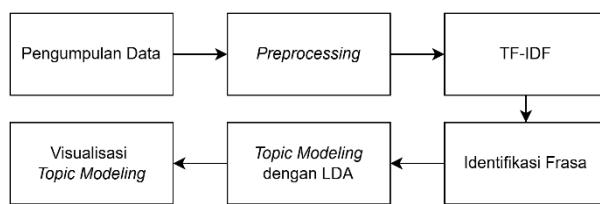
Penelitian ini memiliki kontribusi kebaruan dalam menerapkan metode *topic modeling* LDA secara sistematis pada ulasan konsumen di platform Female Daily. Belum banyak studi sebelumnya yang mengeksplorasi struktur topik laten dalam konteks komunitas kecantikan Indonesia menggunakan pendekatan ini, padahal potensi eksploratif terhadap e-WOM (*electronic word of mouth*) di sektor ini sangat besar. Dengan menyajikan pendekatan berbasis NLP, validasi manusia, dan visualisasi pyLDAvis, penelitian ini memperluas wawasan *text mining* dalam domain kecantikan digital. Oleh karena itu, penelitian ini mengadopsi metode LDA yang bertujuan untuk mengidentifikasi topik utama dalam komentar pengguna terhadap produk *skincare* di Female Daily. Fokus penelitian pada ulasan produk pemenang Female Daily Best of Beauty Awards (BoBA) tahun 2020–2024. Female Daily dipilih sebagai sumber data karena merupakan platform komunitas kecantikan terbesar di Indonesia yang berfokus secara spesifik pada diskusi dan ulasan produk *skincare*. Berbeda dengan platform seperti Shopee, Instagram, atau TikTok yang bersifat komersial atau berbasis konten visual dan promosi singkat, Female Daily menawarkan ulasan yang lebih mendalam, terstruktur, dan berasal dari pengalaman pribadi pengguna.

Metode LDA diterapkan pada penelitian ini untuk *topic modeling* dengan menggunakan nilai *coherence* dalam

menentukan banyaknya topik. Model LDA menghasilkan sejumlah topik yang mencerminkan persepsi pengguna terhadap produk. Visualisasi dilakukan menggunakan pyLDAvis untuk membantu interpretasi distribusi dan keterkaitan antar topik. Hasil penelitian ini diharapkan dapat memberikan wawasan bagi pengembang produk *skincare* dalam memahami kebutuhan dan preferensi konsumen secara lebih mendalam. Selain itu, diharapkan penelitian ini bermanfaat dalam memberikan informasi yang relevan bagi masyarakat luas mengenai produk-produk tersebut.

II. METODE

Penelitian ini dilakukan sesuai dengan tahapan pada Gambar 1. Tahap pertama dilakukan proses pengumpulan data dari komentar produk *skincare* pada Female Daily. Kemudian dilakukan *preprocessing* untuk membersihkan data. Tahap berikutnya yaitu mengukur bobot atau pentingnya suatu kata menggunakan TF-IDF. Kemudian dilanjutkan dengan tahap pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA). Tahap terakhir dilakukan visualisasi dari hasil analisis *topic modeling*.



Gambar 1. Tahapan Penelitian

A. Pengumpulan Data

Proses pengumpulan data dilakukan dengan mengambil data komentar dari produk *skincare* pada forum *online* Female Daily, yaitu dengan memanfaatkan *web scraping*. *Web scraping* merupakan teknik pengambilan data tertentu secara semi-terstruktur pada sebuah halaman web. Komentar yang diambil melalui *scraping* adalah data dari komentar produk *skincare* yang mendapat penghargaan dalam *Best of Beauty Awards* di Female Daily periode tahun 2020 hingga 2024 dengan rentang waktu komentar antara 2015 hingga 2024. Rentang waktu *scraping* dilakukan antara Januari hingga Februari 2025.

B. Preprocessing Data

Tahap *preprocessing* berperan penting dalam pengolahan data sebelum masuk ke tahap berikutnya, memilah data yang perlu digunakan, dan mencegah data yang hilang serta tidak konsisten sehingga dapat menghambat proses analisis dan memberikan hasil yang tidak akurat. Berikut tahapan *preprocessing* dalam penelitian ini:

1. *Remove punctuation* merupakan proses pembersihan elemen non-alfanumerik seperti angka, simbol khusus, dan tanda baca dari korpus teks dianggap tidak membawa informasi penting.

2. *Case folding* merupakan tahapan mengubah seluruh huruf dalam dokumen menjadi huruf kecil (*lowercase*).
3. *Tokenizing* merupakan tahapan untuk memecah teks menjadi unit-unit lebih kecil yang disebut token.
4. *Normalization* merupakan tahapan untuk mengubah kata-kata yang tidak standar pada data komentar menjadi sesuai dengan standar.
5. *Stopwords removal* merupakan tahapan menghapus kata-kata yang dianggap tidak memiliki arti dalam pemrosesan, seperti “dan”, “yang”, “di”, sehingga hanya terfokus pada kata-kata penting yang membawa informasi lebih penting.
6. *Stemming* merupakan proses reduksi kata-kata menjadi bentuk dasarnya.

C. Identifikasi Frasa dengan Bigram dan Trigram

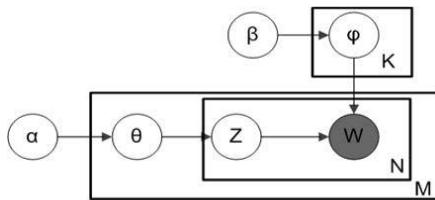
Bigram adalah pasangan dua kata yang sering muncul bersama dalam suatu dokumen, sedangkan trigram adalah rangkaian tiga kata dengan karakteristik serupa. Tahap identifikasi frasa ini bertujuan untuk mengekstraksi frasa kandidat berdasarkan pola kemunculan bersama (*co-occurrence*) kata-kata dalam bentuk bigram dan trigram.

D. TF-IDF

Metode *Term Frequency-Inverse Document Frequency* atau TF-IDF merupakan suatu metrik numerik yang digunakan untuk mengukur bobot atau pentingnya suatu kata (*term*) dalam suatu konten dari kumpulan dokumen. Pengukuran bobot *term* pada sebuah dokumen mencerminkan tingkat kepentingan suatu kata terhadap dokumen, semakin sering suatu kata muncul maka semakin tinggi bobot yang diberikan, sehingga mengindikasikan bahwa kata tersebut memiliki kontribusi yang besar terhadap makna dokumen. Metrik ini mempertimbangkan dua faktor utama yaitu *Term Frequency* (TF) atau frekuensi kemunculan kata yang muncul dalam suatu dokumen dan *Inverse Document Frequency* (IDF) atau frekuensi dokumen yang mengandung kata tersebut [16].

E. Topic modeling menggunakan Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation atau LDA merupakan model probabilistik yang pertama kali diusulkan Blei [12]. LDA merupakan algoritma statistik yang bersifat generatif dan *unsupervised* yang digunakan untuk mengungkap tema-tema laten dari kumpulan dokumen dengan pendekatan *Bayesian* [16]. LDA memanfaatkan distribusi *Dirichlet* untuk menghasilkan distribusi topik pada setiap dokumen. Proses generatif yang dihasilkan oleh distribusi *Dirichlet* kemudian digunakan untuk mendistribusikan kata-kata dalam suatu dokumen ke berbagai topik yang berbeda. Objek yang diamati oleh LDA adalah dokumen; sedangkan distribusi kata dengan topik, distribusi topik dengan dokumen, serta penentuan topik dari setiap kata bersifat tidak terlihat atau tersembunyi (*latent*), maka model ini dinamakan *Latent Dirichlet Allocation*.



Gambar 2. Visualisasi Grafik Model LDA

LDA digambarkan dengan model grafik menggunakan *plate notation* [15] sebagaimana ditunjukkan pada Gambar 2. *Plate notation* merupakan cara visual untuk merepresentasikan variabel atau entitas yang berulang dalam model probabilistik. Variabel dalam *plate notation* untuk LDA sebagai berikut:

- β : *Dirichlet parameter* atas distribusi kata terhadap topik
- φ : Distribusi kata terhadap topik dalam *corpus*
- K : Kumpulan topik
- W : Kata
- N : Kumpulan kata
- M : Kumpulan dokumen
- Z : Topik *index assignment*
- θ : Dokumen
- α : *Dirichlet parameter* atas distribusi topik terhadap dokumen

Struktur grafik pada Gambar 2 menunjukkan bahwa parameter α dan β hanya diambil sampel sekali untuk seluruh korpus dokumen, sedangkan topik untuk setiap dokumen dan kata-kata dalam setiap dokumen diambil sampel secara berulang berdasarkan topik yang telah ditentukan sebelumnya. LDA dapat dihitung menggunakan rumus yang ditunjukkan pada persamaan (1).

$$p(w, z, \theta, \varphi | \alpha, \beta) = p(\varphi | \beta) p(\theta | \alpha) p(z | \theta) p(w | \varphi k) \quad (1)$$

Pada penelitian ini, LDA diterapkan untuk *topic modeling* dengan menggunakan *coherence score* dalam menentukan banyaknya model. Semakin tinggi *coherence score*, semakin baik juga pemahaman interpretasi manusia terhadap topik tersebut. *Coherence score* dihitung menggunakan modul CoherenceModel pada library Gensim. Pada setiap iterasi, model LDA dilatih menggunakan LdaModel dengan *corpus*, *dictionary*, *num_topics*, *random state*, jumlah *iterations*, dan jumlah *passes* yang telah ditentukan untuk mendapatkan topik dengan model yang optimal.

F. Visualisasi Topic Modeling

Hasil dari *topic modeling* divisualisasikan menggunakan pyLDAvis untuk membantu melihat kata-kata yang paling relevan untuk setiap topik dan memahami hubungan antar topik.

III. HASIL DAN PEMBAHASAN

Penelitian ini mengambil data komentar sebanyak 73 produk *skincare* dengan metode *web scraping* dari Female Daily API. Data komentar terkumpul sebanyak 66.521 yang disimpan ke dalam *file* dengan format .xlsx. *Dataset* kemudian diubah menjadi korpus untuk membuat *topic modeling*. *Dataset* dibersihkan terlebih dahulu melalui *preprocessing* dengan urutan tahapan yaitu *remove punctuation*, *case folding*, *tokenizing*, *normalization*, *stopwords removal*, dan *stemming*. Tabel 1 menunjukkan contoh data yang telah melalui tahap *remove punctuation* dan *case folding*.

TABEL 1
DATA KOMENTAR REMOVE PUNCTUATION DAN CASE FOLDING

Komentar	Komentar Setelah Case Folding
Bagus bgt! toner nya bener2 melembabkan kulit aku yg super kering. aku pakai udah 2 minggu dan sekarang kulit aku udh gag kering 😊. wangi nya itu soft tp nyegerin. Bakal beli lagi kalau udh abis. hehe	bagus bgt toner nya bener melembabkan kulit aku yg super kering aku pakai udah minggu dan sekarang kulit aku udh gag kering wanginya itu soft tp nyegerin bakal beli lagi kalau udh abis hehe

Remove punctuation dilakukan untuk menghilangkan tanda baca dan simbol non-alfabetik dari korpus teks. Selanjutnya, *case folding* dilakukan untuk mengubah semua huruf menjadi huruf kecil. Penggabungan penyajian hasil kedua proses ini dimungkinkan karena perubahan data yang dihasilkan oleh *remove punctuation* dan *case folding* tidak memiliki perbedaan yang signifikan secara substansial, sehingga dapat disajikan secara kolektif tanpa mengurangi informasi penting dari tahap pembersihan awal. Tahap selanjutnya adalah *tokenizing*, contoh implementasi dari tahap ini ditunjukkan pada Tabel 2.

TABEL 2
DATA KOMENTAR TOKENIZING

Komentar Setelah Case Folding	Komentar Setelah Tokenizing
bagus bgt toner nya bener melembabkan kulit aku yg super kering aku pakai udah minggu dan sekarang kulit aku udh gag kering wanginya itu soft tp nyegerin bakal beli lagi kalau udh abis hehe	[‘bagus’, ‘bgt’, ‘toner’, ‘nya’, ‘bener’, ‘melembabkan’, ‘kulit’, ‘aku’, ‘yg’, ‘super’, ‘kering’, ‘aku’, ‘pakai’, ‘udah’, ‘minggu’, ‘dan’, ‘sekarang’, ‘kulit’, ‘aku’, ‘udh’, ‘gag’, ‘kering’, ‘wangi’, ‘nya’, ‘itu’, ‘soft’, ‘tp’, ‘nyegerin’, ‘bakal’, ‘beli’, ‘lagi’, ‘kalau’, ‘udh’, ‘abis’, ‘hehe’]

Tokenizing berfungsi untuk memisahkan sebuah korpus teks menjadi unit-unit linguistik individual yang lebih kecil, yang dikenal sebagai token. Selanjutnya, tahap *normalization* dilakukan untuk mengubah kata-kata yang tidak standar pada data komentar menjadi sesuai dengan standar. Hasil dari tahap *normalization* dapat dilihat pada Tabel 3.

TABEL 3
DATA KOMENTAR NORMALIZATION

Komentar Setelah Tokenizing	Komentar Setelah Normalization
[‘bagus’, ‘bgt’, ‘toner’, ‘nya’, ‘bener’, ‘melembabkan’, ‘kulit’, ‘aku’, ‘yg’, ‘super’, ‘kering’, ‘aku’, ‘pakai’, ‘udah’, ‘minggu’, ‘dan’, ‘sekarang’, ‘kulit’, ‘aku’, ‘udh’, ‘gag’, ‘kering’, ‘wangi’, ‘nya’, ‘itu’, ‘soft’, ‘tp’, ‘nyegerin’, ‘bakal’, ‘beli’, ‘lagi’, ‘kalau’, ‘udh’, ‘abis’, ‘hehe’]	bagus banget toner nya benar melembabkan kulit aku yang super kering pakai sudah minggu dan sekarang enggak wanginya itu soft tapi nyegerin bakal beli lagi kalau habis hehe

Tahap selanjutnya dilakukan *stopwords removal* untuk menghapus kata-kata yang dianggap tidak memiliki arti dalam pemrosesan sehingga hanya terfokus pada kata-kata penting yang membawa informasi lebih penting. Hasil dari tahap ini ditunjukkan pada Tabel 4.

TABEL 4
DATA KOMENTAR STOPWORDS REMOVAL

Komentar Setelah Normalization	Komentar Setelah Stopwords Removal
bagus banget toner nya benar melembabkan kulit aku yang super kering pakai sudah minggu dan sekarang enggak wanginya itu soft tapi nyegerin bakal beli lagi kalau habis hehe	toner melembabkan kulit kering

Tahap terakhir dilakukan *stemming* untuk mereduksi kata-kata menjadi bentuk dasarnya. Hasil dari tahap ini dapat dilihat pada Tabel 5.

TABEL 5
DATA KOMENTAR STEMMING

Komentar Setelah Stopwords Removal	Komentar Setelah Tokenizing
toner melembabkan kulit kering	toner lembab kulit kering

Setelah melalui proses *preprocessing*, jumlah token yang diperoleh dari seluruh dokumen ulasan adalah sebanyak 582.492 token. Token tersebut selanjutnya dijadikan *dataset* untuk dilakukan pembuatan model bigram dan trigram. Proses yang dilakukan mencakup pencarian dan penggabungan kata-kata yang berdekatan menjadi unit frasa, sesuai dengan pola bigram dan trigram. Tahap ini menggunakan modul *phrases* dan *phraser* dari *library* gensim pada Python dengan ketentuan parameternya dapat diliat pada Gambar 3.

```
# model Bigram
bigram = Phrases(sentences, min_count=10,
                  threshold=20, delimiter='_')
bigram_model = Phraser(bigram)

# model Trigram
trigram = Phrases(bigram_model[sentences],
                  min_count=10, threshold=20,
```

```
delimiter='_')
trigram_model = Phraser(trigram)
```

Gambar 3. Kode Program Model Bigram dan Trigram

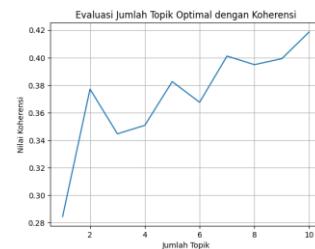
Model bigram dibuat dengan parameter `min_count=10` dan `threshold=20`. Parameter `min_count=10` menetapkan bahwa suatu pasangan kata hanya akan diproses jika muncul minimal sebanyak 10 kali, sedangkan `threshold=20` berfungsi sebagai ambang nilai minimum yang harus dicapai agar pasangan kata tersebut dapat dikenali sebagai suatu frasa. Penggabungan frasa menggunakan tanda hubung *underscore*. Hasil dari identifikasi frasa dapat dilihat pada Tabel 6.

TABEL 6
HASIL IDENTIFIKASI FRASA

Komentar Setelah Preprocessing	Data Setelah Bigram dan Trigram
glowing bekas jerawat pudar packaging tekstur serum cair	glowing bekas_jerawat pudar packaging tekstur serum cair

Selanjutnya, dilakukan pembobotan dengan TF-IDF. Pembobotan menggunakan fungsi dari `doc2bow` untuk mengonversi teks menjadi representasi TF (*Term Frequency*). Data masukan akan dipecah menjadi kata-kata (tokenisasi), kemudian dihitung frekuensi kemunculan setiap kata dalam dokumen.

Data yang telah dilakukan pembobotan selanjutnya dilakukan analisis dengan *topic modeling* untuk mengidentifikasi topik-topik yang dibicarakan pada suatu dokumen atau data yang telah dikumpulkan. Metode *Latent Dirichlet Allocation* (LDA) diterapkan pada penelitian ini untuk *topic modeling* dengan menggunakan *coherence score* dalam menentukan banyaknya model. *Coherence score* berfungsi sebagai metrik untuk menilai seberapa mudah kumpulan kata yang dihasilkan oleh model topik dapat dipahami secara manusiawi [18]. Semakin tinggi *coherence value*, semakin baik pula interpretasi manusia [19]. Setiap model LDA dibangun menggunakan *library* `LdaModel` dari Gensim, dan nilai koherensinya dihitung menggunakan `CoherenceModel` dengan metrik `c_v`.



Gambar 4. Grafik Coherence Score

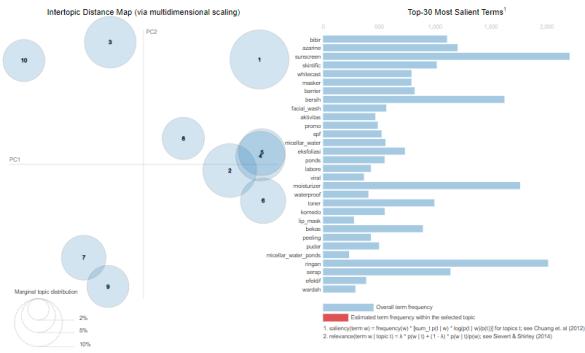
Gambar 4 merupakan visualisasi grafik *coherence score* menunjukkan nilai tertinggi pada 10 topik setelah pengujian dengan variasi jumlah topik dari 1 hingga 10. Nilai *coherence score* tertinggi ini adalah 0,418738, dengan detail yang tertera pada Tabel 7. Oleh karena itu, sebanyak 10 topik ditetapkan sebagai parameter `num_topic` pada model LDA, karena

dianggap menghasilkan representasi semantik paling koheren dari data komentar *skincare* Female Daily.

TABEL 7
COHERENCE VALUE

Topik ke-	Nilai Coherence
1	0,284595
2	0,377177
3	0,344661
4	0,350839
5	0,382594
6	0,367590
7	0,401182
8	0,394930
9	0,399275
10	0,418738

Kemudian dilakukan analisis dengan melihat visualisasi dari *tools* pyLDAvis serta hubungan antar *term* dalam suatu topik, dalam penelitian ini komentar produk *skincare* pada Female Daily. Visualisasi pyLDAvis menyajikan dua komponen utama yang saling terhubung, seperti terlihat pada Gambar 5. Pada bagian kiri visualisasi, topik-topik direpresentasikan dalam bentuk gelembung, di mana ukuran gelembung merepresentasikan frekuensi kemunculan topik pada data komentar *skincare* Female Daily. Posisi dan kedekatan antar gelembung mencerminkan relasi atau korelasi antar topik yang ditentukan berdasarkan nilai bobot. Gelembung yang tersebar atau tidak saling beririsan mengindikasikan minimnya keterkaitan antar topik, sementara gelembung yang berdekatan atau beririsan menunjukkan sebaliknya.

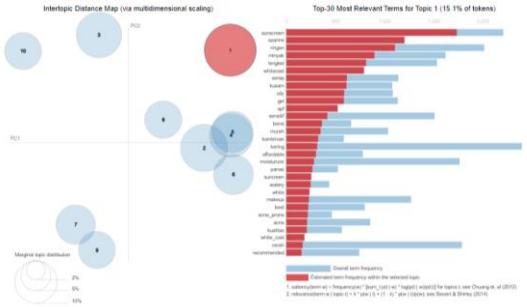


Gambar 5. Visualisasi LDA Komentar Produk Skincare pada Female Daily

Bar chart pada bagian kanan visualisasi menampilkan frekuensi 30 kata yang sering dibicarakan pada topik. Terdapat juga diagram batang horizontal yang merepresentasikan kata-kata utama dalam masing-masing topik. Panjang setiap batang menunjukkan seberapa sering kata tersebut muncul dalam topik. Visualisasi ini menghubungkan bagian kiri dan kanan untuk memperlihatkan relasi antara distribusi kata dan representasi topik. Visualisasi

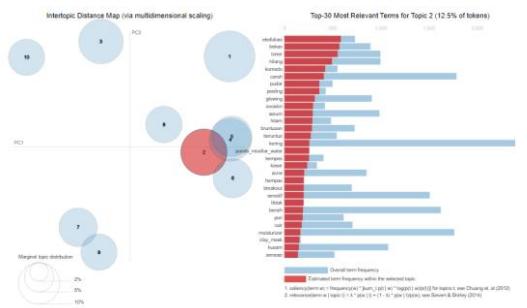
topik yang disajikan pada Gambar 5 menunjukkan adanya gelembung yang beririsan, yaitu:

- *Topic 2* beririsan dengan *topic 4, 5, dan 6*
- *Topic 4* beririsan dengan *topic 2, 5, dan 6*
- *Topic 5* beririsan dengan *topic 2 dan 4*
- *Topic 6* beririsan dengan *topic 2 dan 4*
- *Topic 7* beririsan dengan *topic 9*
- *Topic 9* beririsan dengan *topic 7*



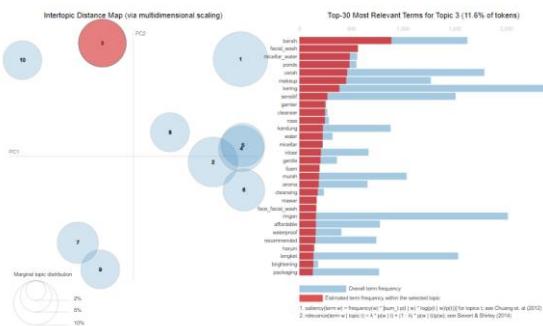
Gambar 6. Visualisasi topic ke-1

Gambar 6 menampilkan kata-kata yang muncul dalam topik ke-1. Kata-kata dengan probabilitas tertinggi yang mewakili *topic* ke-1 yaitu: $0,053^*\text{masker} + 0,019^*\text{pori} + 0,014^*\text{mugwort} + 0,012^*\text{residu} + 0,011^*\text{bersih} + 0,010^*\text{masker_mugwort} + 0,009^*\text{acne} + 0,009^*\text{kering} + 0,009^*\text{cerah} + 0,008^*\text{whitelab}$. Pada topik ke-1, sepuluh *term* teratas tersebut menunjukkan bahwa komentar produk *skincare* cenderung membahas terkait masker wajah merek Whitelab dengan kandungan mugwort. Lebih lanjut, pembahasan tersebut terkait terkait dengan aspek manfaat produk dalam menangani permasalahan kulit wajah yang berkaitan dengan pori-pori, jerawat, dan kulit berminyak.



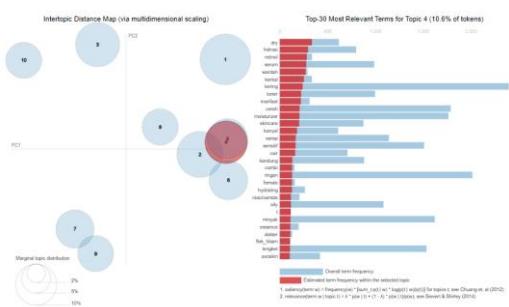
Gambar 7. Visualisasi topic ke-2

Gambar 7 menampilkan kata-kata yang muncul dalam topik ke-2. Kata-kata dengan probabilitas tertinggi yang mewakili *topic* ke-2 yaitu: $0,058^*\text{skintific} + 0,043^*\text{barrier} + 0,041^*\text{moisturizer} + 0,025^*\text{viral} + 0,013^*\text{kering} + 0,011^*\text{sensitif} + 0,010^*\text{ringan} + 0,010^*\text{skin_barrier} + 0,009^*\text{clay} + 0,008^*\text{skincare}$. Pada topik ke-2, *term* teratas mengindikasikan adanya fokus pembahasan pada produk *moisturizer* dari merek Skintific yang berkaitan dengan kulit kering, terutama untuk perawatan *skin barrier*.



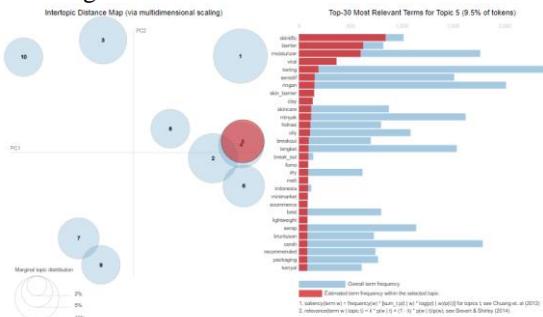
Gambar 8. Visualisasi *topic* ke-3

Gambar 8 menampilkan kata-kata yang muncul dalam topik ke-3. Kata-kata dengan probabilitas tertinggi yang mewakili topik ke-3 yaitu: '0,050*"bersih" + 0,032*"facial_wash" + 0,028*"micellar_water" + 0,027*"ponds" + 0,026*"cerah" + 0,026*"makeup" + 0,022*"kering" + 0,015*"sensitif" + 0,014*"garnier" + 0,014*"cleanser"'. Pada topik ke-3, sepuluh *term* teratas mengindikasikan adanya fokus pembahasan pada produk *skincare* pembersih wajah dan meyoroti merek Pond's.



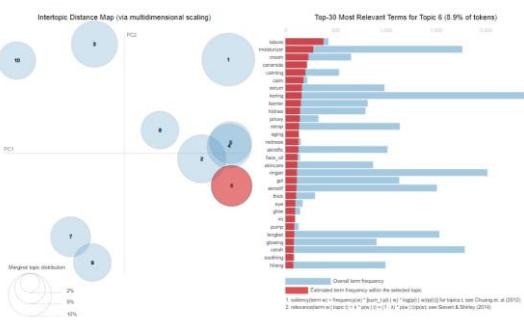
Gambar 9. Visualisasi *topic* ke-4

Gambar 9 menampilkan kata-kata yang muncul dalam topik ke-4. Kata-kata dengan probabilitas tertinggi yang mewakili topik ke-4 yaitu: '0.080*"bibir" + 0,040*"kering" + 0,021*"lip_mask" + 0,017*"awet" + 0,015*"ringan_white_cast" + 0,014*"lip" + 0,012*"laneige" + 0,012*"cerah" + 0,011*"murah" + 0,011*"emina"'. Pada topik ke-4, sepuluh *term* teratas tersebut berdiskusi produk perawatan bibir, terutama produk *lip mask* untuk bibir kering dan manfaatnya untuk menjadikan bibir tampak lebih cerah. Merek-merek populer dalam topik ini yang turut muncul adalah Laneige dan Emina.



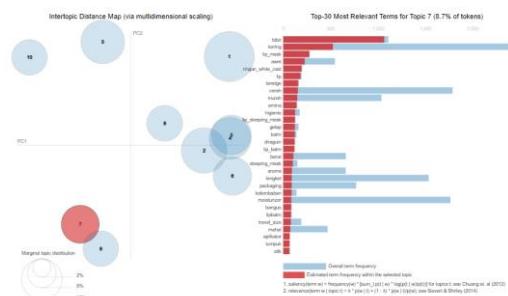
Gambar 10. Visualisasi *topic* ke-5

Gambar 10 menampilkan kata-kata yang muncul dalam topik ke-5. Kata-kata dengan probabilitas tertinggi yang mewakili topik ke-5 yaitu: '0,017*"praktis" + 0,010*"skin_preparation" + 0,010*"lengket_white_cast" + 0,009*"moisture" + 0,009*"ringan" + 0,009*"hyaluronic_acid" + 0,009*"cooling" + 0,008*"iklan" + 0,008*"kandung" + 0,008*"makeup"'. Pada topik ke-10, *term* teratas tersebut menunjukkan karakteristik yang berbeda dari topik-topik sebelumnya dan sulit untuk diinterpretasikan secara koheren. Kumpulan *term* ini mencakup konsep-konsep yang berasal dari ranah semantik yang sangat beragam dan tidak menunjukkan keterikatan tematik yang kuat satu sama lain. Kata "kandung" dan "hyaluronic acid" merepresentasikan aspek kandungan dalam produk *skincare*. Kata "skin preparation" secara khusus merujuk pada rutinitas pra-aplikasi *makeup*. Pengalaman pengguna terwujud dalam kata-kata seperti "lengket," "white cast," "moisture," "cooling," "praktis," dan "ringan." Kategori produk *makeup* berbeda dari *skincare* dan umumnya berdiri sendiri dalam segmentasi industri kecantikan. Selanjutnya, istilah "iklan" terkait erat dengan aktivitas pemasaran.



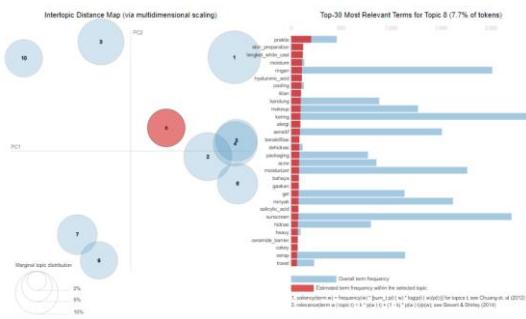
Gambar 11. Visualisasi *topic* ke-6

Gambar 11 menampilkan kata-kata yang muncul dalam topik ke-6. Kata-kata dengan probabilitas tertinggi yang mewakili topik ke-6 yaitu: "0,075*"sunscreen" + 0,052*"azarine" + 0,048*"ringan" + 0,039*"minyak" + 0,035*"lengket" + 0,034*"whitecast" + 0,027*"serap" + 0,027*"kusam" + 0,025*"oily" + 0,025*"gel"'. Pada topik ke-6, sepuluh *term* teratas tersebut terlihat adanya fokus pembahasan mengenai produk *sunscreen* dari merek Azarine dan membicarakan tentang kenyamanan produk yang ringan, cepat menyerap, dan tidak meninggalkan *whitecast*.



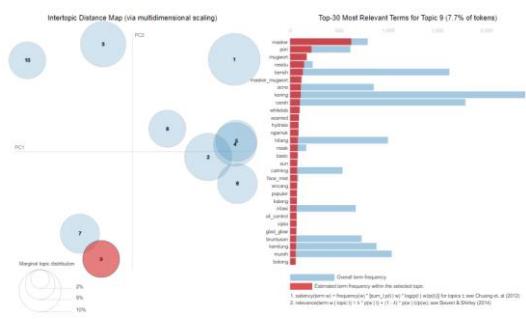
Gambar 12. Visualisasi topic ke-7

Gambar 12 menampilkan kata-kata yang muncul dalam topik ke-7. Kata-kata dengan probabilitas tertinggi yang mewakili topik ke-7 yaitu: ‘0,031*“eksfoliasi” + 0,030*“bekas” + 0,029*“toner” + 0,026*“hilang” + 0,022*“komedo” + 0,021*“cerah” + 0,019*“pudar” + 0,019*“peeling” + 0,017*“glowing” + 0,016*“avoskin”’. Pada topik ke-7, sepuluh *term* teratas tersebut membahas mengenai produk toner eksfoliasi dari merek Avoskin untuk masalah kulit seperti bekas jerawat dan komedo dengan manfaatnya untuk mencapai hasil seperti bekas yang pudar atau hilang, kulit yang tampak lebih cerah, dan memberikan efek *glowing*.



Gambar 13. Visualisasi topic ke-8

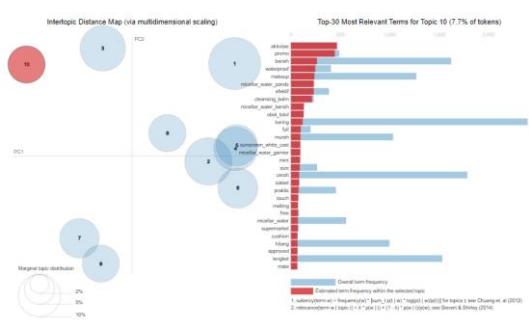
Gambar 13 menampilkan kata-kata yang muncul dalam topik ke-8. Kata-kata dengan probabilitas tertinggi yang mewakili topik ke-8 yaitu: ‘0,021*“dry” + 0,018*“hidrasi” + 0,017*“retinol” + 0,017*“serum” + 0,017*“wardah” + 0,016*“kental” + 0,015*“kering” + 0,014*“toner” + 0,014*“manfaat” + 0,013*“cerah”’. Pada topik ke-8, sepuluh *term* teratas tersebut membahas toner dan serum dengan kandungan aktif seperti retinol yang cocok untuk kulit kering, dengan manfaat yang diberikan untuk kulit cerah dan lembap serta menyoroti merek Wardah.

TABEL 8
ANALISIS TOPIK BERIRISAN

Topik Beririsan	Kata Kunci Dominan	Penjelasan
2 dan 4	Topik 2: moisturizer, barrier, sensitif Topik 4: bibir, kering, lip_mask	Kedua topik menyoroti isu kelembapan dan perlindungan kulit. Topik 2 fokus pada pelembab wajah, sedangkan Topik 4 pada produk bibir. Irisan terjadi karena keduanya sama-sama membahas produk yang digunakan untuk kulit kering dan sensitif.

Gambar 14. Visualisasi topic ke-9

Gambar 14 menampilkan kata-kata yang muncul dalam topik ke-9. Kata-kata dengan probabilitas tertinggi yang mewakili topik ke-9 yaitu: ‘0,028*“labore” + 0,020*“moisturizer” + 0,017*“cream” + 0,016*“ceramide” + 0,015*“calming” + 0,013*“calm” + 0,012*“serum” + 0,012*“kering” + 0,011*“barrier” + 0,011*“hidrasi”’. Pada topik ke-9, sepuluh *term* teratas tersebut membahas mengenai *moisturizer* berbentuk *cream* dari merek Labore yang memiliki kandungan ceramide sebagai solusi multifungsi untuk kulit kering, tidak hanya untuk hidrasi tetapi juga perbaikan *skin barrier* dan efek menenangkan.



Gambar 15. Visualisasi topic ke-10

Gambar 15 menampilkan kata-kata yang muncul dalam topik ke-10. Kata-kata dengan probabilitas tertinggi yang mewakili topik ke-10 yaitu: ‘0,040*“aktivitas” + 0,038*“promo” + 0,023*“bersih” + 0,021*“waterproof” + 0,020*“makeup” + 0,020*“micellar_water_ponds” + 0,020*“efektif” + 0,018*“cleansing_balm” + 0,011*“micellar_water_bersih” + 0,011*“obat_totol”’. Pada topik ke-10, sepuluh *term* teratas tersebut mengindikasikan pembahasan mengenai produk pembersih wajah yang efektif untuk membersihkan kotoran wajah setelah aktivitas dan *makeup* yang *waterproof*.

Untuk memahami struktur topik secara lebih dalam maka dilakukan analisis irisan antar topik. Irisan ini mengindikasikan adanya kemiripan atau tumpang tindih antara beberapa topik. Analisis ini bertujuan untuk memastikan kebenaran bahwa visualisasi yang dihasilkan telah merepresentasikan hubungan topik secara akurat dan valid berdasarkan data.

2 dan 5	Topik 2: moisturizer, barrier, sensitif Topik 5: ringan, moisture, hyaluronic_acid	Produk pada kedua topik dibahas dalam konteks kenyamanan tekstur dan hidrasi, dengan kata kunci seperti "ringan" dan "moisture" menjadi penghubung tematik.
2 dan 6	Topik 2: moisturizer, barrier, sensitif Topik 6: sunscreen, ringan, whitecast	Moisturizer dan sunscreen dinilai berdasarkan kesamaan atribut, seperti ringan, tidak lengket, dan cocok untuk kulit sensitif. Pengguna kerap membandingkan kenyamanan pemakaian antara keduanya.
5 dan 4	Topik 5: moisture, hyaluronic acid, ringan Topik 4: bibir, kering, lip_mask	Irisan terjadi karena keduanya sama-sama membahas aspek hidrasi dan kenyamanan pemakaian. Lip mask dan pelembab wajah dibahas dalam konteks tekstur dan sensasi ringan di kulit.
6 dan 4	Topik 6: sunscreen, ringan, whitecast Topik 4: bibir, kering, ringan_white_cast	Kedua topik sama-sama membahas produk yang dinilai dari pengalaman sensorik seperti "ringan" dan tidak meninggalkan "whitecast". Ini menunjukkan bahwa kenyamanan kosmetik menjadi perhatian utama pengguna.
7 dan 9	Topik 7: eksfoliasi, bekas, cerah Topik 9: moisturizer, ceramide, calming	Topik 7 menggambarkan tindakan perawatan aktif (eksfoliasi), sedangkan Topik 9 membahas pemulihan dan proteksi kulit. Irisan timbul karena kedua topik merupakan tahapan berurutan dalam rutinitas perawatan kulit.

Analisis ini menunjukkan bahwa topik-topik yang dihasilkan saling terkait dan tidak berdiri sendiri. Keterkaitan semantik ini mencerminkan pola berpikir pengguna yang menyatukan manfaat, tekstur, efek, dan rutinitas pemakaian produk skincare secara menyeluruh.

Untuk memastikan bahwa topik-topik yang dihasilkan oleh model LDA relevan secara semantik dan sesuai dengan diskusi nyata dalam komunitas pengguna *skincare*, penelitian ini menerapkan metode validasi berbasis penilaian manusia (*human judgment*). Pendekatan ini bertujuan untuk mengevaluasi apakah kumpulan kata kunci dalam setiap topik dapat dipahami dan diinterpretasikan secara logis. Validasi ini menggunakan pendekatan penilaian manusia oleh dua evaluator:

- Evaluator 1: *Beauty influencer* yang aktif dalam media sosial Instagram dan pengguna aktif Female Daily,

- Evaluator 2: Peneliti sendiri sebagai pengguna aktif Female Daily dan sebagai *beauty enthusiast*.

Sebanyak 10 topik dari model LDA terbaik dipilih berdasarkan nilai *coherence* tertinggi (0,418738). Setiap topik diekstraksi kata kunci dominannya dan disertai satu sampel komentar sebagai konteks dokumen. Penilaian relevansi dengan skala:

- Sangat Relevan,
- Relevan,
- Cukup Relevan,
- Tidak Relevan.

Suatu topik dikategorikan "Valid" apabila memperoleh nilai "Relevan" atau lebih tinggi dari setidaknya satu evaluator. Validasi ini mengacu pada pendekatan dalam studi [9] yang menunjukkan pentingnya validasi manusia dalam mengevaluasi keterbacaan dan keterwakilan makna dari topik LDA yang bersifat *unsupervised*.

TABEL 9
HASIL VALIDASI EVALUATOR

Topik	Top-3 Kata Kunci	Sampel Komentar	Skor E1	Skor E2	Status
1	masker, pori, mugwort	"ini masker mugwort pertama aku dan aku langsung suka. cocok banget untuk kondisi kulit aku yg sering muncul jerawat, dan pori2 gede. cukup pake ini 2-3 seminggu dan cukup didiamkan 10-15 menit ajaa. recommended deh"	1	1	Valid
2	skintific, barrier, moisturizer	"Waktu pertama kali menggunakan Moisturaizer Skintific Barrier ini. Sangat amat terbukti jelas prosesnya. Kulit terasa lembab dan ternyata untuk tipe wajah oilypun cocok. Merah di wajah hilang dan tekstur pada wajah lebih halus. WAJIB BANGET COBA !!! Saking cintanya saat ini pakai yang BIGSIZE loh 🤗"	1	1	Valid
3	bersih, facial_wash, ponds	"Dari dulu sampe sekarang ganti kemasan, selalu suka pake facial wash ponds ini. karena menurutku ini cocok banget sama kulitku ini. apalgi bikin kulit tambah cerah, selain facial wash ponds aku juga pake pelembabnya juga, agar kulitku jadi lembab dan ternutrisi."	2	1	Valid
4	bibir, kering, lip_mask	"sudah menggunakan lip mask ini selama lebih dari 1tahun. so far produknya oke, mengatasi bibir kering dan pecah-pecah. walaupun kemasannya kecil tapi awet banget, udah setahun lebih tapi masih belum habis. sudah pasti akan repurchase lagi."	1	1	Valid

5	praktis, <i>skin_preparation</i> , lengket_white_cast	"ini enak banget di kulit terasa lembab spf-nya juga oke teksturnya juga lembut sih di wajah aku nggak ninggalin white cast kayak pakai pelembab aja tapi kulit tetap dilindungi dong sama sunscreennya packaging-nya juga oke praktis ringan dibawa ke mana-mana"	3	3	Valid
6	<i>sunscreen</i> , azarine, ringan	"sunscreen azarine ringan bangett teksturnya terus cepet nyerep, gak peeling dan gak bikin minyakan. cocok buat semua tipe kulit mau yang berminyak maupun kering plus harganya murahhhh dpt banyakkk sangat sangat sukaaaaaaa♡♡♡♡@@@♡♡♡"	1	1	Valid
7	eksfoliasi, bekas, <i>toner</i>	"Jatuh cinta sama ysb lactid acid avoskin ini buat eksfoliasi malam hari.. baru seminggu pake udah keliatan efek cerahnya walaupun awal pake agak sedikit cekat cekit sementara dibagian jerawat dan paginya muka keliatan glowing. berharap penuh sama ysb ini buat atasi beruntusanku"	3	2	Valid
8	<i>dry</i> , hidrasi, retinol	"Serum Retinol yang cocok untuk pemula, sejauh pemakaian aku produk ini emang bagus banget untuk kulit ku yang kurang terhidrasi, berjerawat dan kusam. Muka langsung lembab, cerah dan lembut. Jerawat juga lebih calm. Harganya juga terjangkau loh. Bisa coba yg ukuran kecil dulu yaa."	2	3	Valid
9	labore, <i>moisturizer</i> , cream	"Luvvv sm yg satu ini! Teksturnya cream gel tapi soft dan nggak thick gitu. Jadi pas dipake ringan banget, nggak kerasa berat & greasy di kulit. Finish moisturizer nya juga bukan yg bikin berminyak atau jadi matte gitu. Lebih ke satin finish dan bikin berasa healthy skin, karna emang melembapkan bgt kulit yg lagi kering♥ this one jadi favorit aku banget sih diantara produk LABORE lainnya!!@@"	1	1	Valid
10	aktivitas, promo, bersih	Bener-bener gampang banget bersihinya cuma pake dikit mascara waterproof langsung meleleh jadi irit makinnya ga perih sama sekali. Gampang dicari offline atau online, sering promo juga. cinta banget deh pokoknya ♡	3	3	Valid

Hasil validasi semantik menunjukkan bahwa keseluruhan topik yang dihasilkan oleh model LDA dapat diinterpretasikan dengan baik oleh evaluator dari kalangan pengguna domain. Dari sepuluh topik yang dievaluasi, enam topik (Topik 1, 2, 4, 6, 9, dan sebagian Topik 3) atau sekitar 60%, memperoleh skor "Sangat Relevan" dari kedua evaluator, sedangkan 30% lainnya (Topik 5, 7, dan 8) dinilai "Relevan" hingga "Cukup Relevan". Satu topik (10%), yaitu Topik 10, menerima penilaian "Cukup Relevan" dari kedua evaluator, tetapi tetap dikategorikan valid karena topik tersebut merepresentasikan praktik nyata seperti efektivitas produk pembersih dan promo. Tidak ada topik yang dinilai "Tidak Relevan". Secara keseluruhan, 100% topik dinilai valid oleh minimal satu evaluator, yang berarti hasil pemodelan tidak hanya kuat secara kuantitatif melalui nilai *coherence*, tetapi juga terbukti relevan secara semantik. Validasi ini mengacu pada pendekatan dalam studi [9] yang menekankan pentingnya peran penilaian manusia untuk menjembatani keterbatasan model *unsupervised* dalam memahami struktur tematik secara natural.

Penelitian ini jika dibandingkan dengan studi [20] yang menggunakan LDA pada ulasan aplikasi Shopee, ditemukan bahwa topik-topik dominan pada penelitian tersebut terkait promosi dan fitur layanan. Demikian pula, [21] menemukan topik *brand promotion* dan *engagement* dalam *tweet* mengenai *skincare*. Perbedaan ini menunjukkan bahwa ulasan di Female Daily lebih menitikberatkan pada pengalaman personal, efektivitas produk, dan kualitas, sejalan dengan karakteristik komunitas *skincare* yang reflektif dan berbasis pengalaman pengguna.

Topik-topik dari penelitian ini, yang dihasilkan dari ulasan pada Female Daily, memberikan tema mengenai efektivitas produk, kecocokan kulit, aroma, tekstur, dan harga. Selain itu, keunggulan penelitian ini terletak pada cakupan data yang jauh lebih luas, yaitu mencakup lebih dari 66.000 komentar, serta fokus spesifik pada produk-produk pemenang Female Daily *Best of Beauty Awards* (BoBA) selama lima tahun terakhir. Juga, berbeda dari studi sebelumnya yang mengandalkan platform media sosial umum, penelitian ini menyajikan pemetaan topik yang lebih representatif dari komunitas kecantikan lokal Indonesia.

Hasil penelitian ini juga sejalan dengan teori *electronic word of mouth* (e-WOM) yang menyatakan bahwa komunikasi daring antar konsumen menjadi faktor penting dalam memengaruhi keputusan pembelian [22]. Temuan topik seperti "pengalaman pribadi", "efek samping", "efektivitas pemakaian", dan "rekomendasi" mencerminkan proses berbagi pengalaman yang merupakan inti dari e-WOM. Berdasarkan penelitian oleh [23], e-WOM memiliki pengaruh signifikan terhadap keputusan pembelian *skincare* MS Glow di kalangan generasi Y, di mana media sosial berperan sebagai saluran utama untuk membentuk persepsi konsumen. Hal ini mendukung bahwa informasi dari pengguna lain di platform seperti Female Daily memiliki kredibilitas tinggi dalam membentuk preferensi dan minat beli.

Hasil topik-topik yang muncul, seperti pentingnya tekstur, kenyamanan, dan kandungan aktif, mencerminkan preferensi konsumen dalam proses pengambilan keputusan yang dipengaruhi oleh e-WOM. Hal ini konsisten dengan teori

perilaku konsumen, bahwa persepsi dan testimoni orang lain secara *online* memengaruhi minat beli dan loyalitas pengguna terhadap produk *skincare*. Dengan demikian, pendekatan ini tidak hanya memperkuat bukti bahwa LDA dapat menangkap preferensi konsumen secara tematis, tetapi juga memberikan nilai tambah dalam memahami pola komunikasi e-WOM berbasis komunitas yang khas dari platform Female Daily.

IV. KESIMPULAN

Penelitian ini bertujuan untuk memetakan topik-topik utama dalam komentar pengguna terhadap produk *skincare* pemenang penghargaan Female Daily *Best of Beauty Awards* (BoBA) dengan menggunakan metode *Latent Dirichlet Allocation* (LDA). Melalui tahapan pengumpulan data, praproses teks, dan pemodelan topik, diperoleh sejumlah topik yang mencerminkan perhatian dan pengalaman pengguna terhadap berbagai aspek produk, seperti efektivitas, kecocokan dengan jenis kulit, harga, kemasan, dan aroma.

Hasil pemodelan menunjukkan bahwa pengguna Female Daily secara aktif membagikan ulasan berdasarkan pengalaman pribadi, terutama menyangkut dampak produk terhadap kulit mereka. Topik-topik yang dihasilkan juga mengindikasikan pentingnya aspek kepercayaan dan harapan konsumen terhadap produk *skincare* yang direkomendasikan komunitas. Temuan ini menegaskan bahwa ulasan pengguna merupakan sumber *insight* penting bagi pelaku industri kecantikan untuk memahami kebutuhan konsumen. Dengan demikian, penerapan LDA dalam konteks ini terbukti efektif untuk mengidentifikasi tema-tema utama dalam ulasan pengguna secara otomatis dan sistematis. Ke depan, hasil penelitian ini dapat menjadi landasan awal bagi pengembangan sistem rekomendasi produk atau analisis sentimen berbasis topik dalam sektor kecantikan di Indonesia.

Penentuan model terbaik dilakukan menggunakan nilai *coherence* dengan hasil pengujian menunjukkan bahwa model dengan topik sebanyak 10 menghasilkan nilai optimal dengan nilai *coherence* sebesar 0,418738. Hasil tersebut diraih dengan melakukan *running* sebanyak 20 iterasi. Selain itu, visualisasi menggunakan pyLDAvis berhasil menyajikan representasi visual dari distribusi topik yang dominan dengan memperlihatkan sebaran kata-kata kunci yang saling berkaitan dalam setiap topik mengenai produk *skincare*.

Secara keseluruhan, hasil penelitian ini dapat memberikan kontribusi kepada perusahaan dalam memahami persepsi konsumen terhadap produk-produk *skincare* berbasis ulasan *online*. Pendekatan yang digunakan dalam penelitian ini bersifat fleksibel dan dapat direplikasi pada sektor lain yang memiliki karakteristik opini berbasis teks, seperti ulasan produk *fashion*, makanan, atau layanan digital. Selain itu, hasil *topic modeling* ini dapat dikembangkan lebih lanjut melalui integrasi dengan analisis sentimen untuk melihat sentimen dominan dalam setiap topik, atau dengan segmentasi demografis untuk mengungkap preferensi pengguna berdasarkan usia, jenis kulit, atau kebiasaan pembelian.

Penelitian ini juga memberikan kontribusi terhadap pengembangan studi *text mining* di domain kecantikan, khususnya dalam konteks komunitas digital lokal. Penerapan metode *unsupervised topic modeling* memungkinkan eksplorasi struktur tematik secara alami dari opini konsumen, melampaui pendekatan klasifikasi sentimen yang cenderung terbatas pada kategorisasi emosi. Temuan ini memperluas pemahaman tentang bagaimana persepsi pengguna terbentuk dalam ruang diskusi daring dan dapat menjadi fondasi untuk strategi *digital branding* dan inovasi produk yang lebih responsif terhadap kebutuhan pasar.

Di samping itu, masyarakat juga dapat mengakses informasi yang relevan mengenai produk-produk *skincare*. Dalam penelitian ini, jumlah data yang dianalisis mencapai 66.521 komentar menjadikan pemrosesan cukup kompleks dan memakan waktu. Ke depan, penelitian lanjutan disarankan untuk mengadopsi metode otomatisasi lanjutan, seperti *deep learning-based text preprocessing*, guna mengatasi kompleksitas dan skala data yang besar.

DAFTAR PUSTAKA

- [1] Kementerian Perindustrian Indonesia, “Rencana Induk Pembangunan Industri Nasional 2015 - 2035,” 2015. Accessed: Apr. 29, 2025. [Online]. Available: <https://kemenperin.go.id/ripin.pdf>
- [2] D. Waluyo, “Fenomena Cantik Industri Kosmetik.” Accessed: May 12, 2024. [Online]. Available: <https://indonesia.go.id/kategori/editorial/7804/fenomena-cantik-industri-kosmetik?lang=1>
- [3] M. Ferdinand and W. S. Ciptono, “Indonesia’s Cosmetics Industry Attractiveness, Competitiveness and Critical Success Factor Analysis,” *Jurnal Manajemen Teori dan Terapan / Journal of Theory and Applied Management*, vol. 15, no. 2, pp. 209–223, Aug. 2022, doi: 10.20473/jmtt.v15i2.37451.
- [4] N. Khotimah *et al.*, “Pengaruh Citra Merek, Kualitas Produk, Online Customer Review, Dan Online Customer Rating Terhadap Keputusan Pembelian Produk Skincare Madame Gie (Studi Kasus Pada Mahasiswa Prodi Manajemen Universitas Bhayangkara Jakarta Raya Angkatan 2020),” 2024.
- [5] A. Elwälde, K. Lü, and M. Ali, “Perceived Derived Attributes of Online Customer Reviews,” *Comput Human Behav*, vol. 56, pp. 306–319, Mar. 2016, doi: 10.1016/j.chb.2015.11.051.
- [6] A. Thoungrungrue, “The Influence of Social Media Intensity and EWOM on Conspicuous Consumption,” *Procedia Soc Behav Sci*, vol. 148, pp. 7–15, Aug. 2014, doi: 10.1016/j.sbspro.2014.07.009.
- [7] Regita Wahyu and Dian Widya Putri, “Pengaruh Electronic Word of Mouth Website Female Daily terhadap Keputusan Followers Membeli Produk Kosmetik,” *Bandung Conference Series: Public Relations*, vol. 3, no. 2, pp. 941–948, Aug. 2023, doi: 10.29313/bcpr.v3i2.9417.
- [8] Female Daily, “Female Daily: About Us.” Accessed: May 12, 2024. [Online]. Available: <https://femalesdaily.com/about>
- [9] E. Puspita, D. F. Shiddiq, and F. F. Roji, “Pemodelan Topik pada Media Berita Online Menggunakan Latent Dirichlet Allocation (Studi Kasus Merek Somethinc),” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 481–489, Feb. 2024, doi: 10.57152/malcom.v4i2.1204.
- [10] M. Mustahidah, “Topic Modeling pada Ulasan Hotel Menggunakan Latent Dirichlet Allocation (LDA) dan Probabilistic Latent Semantic Analysis (PLSA),” 2021. Accessed: Oct. 18, 2024. [Online]. Available: <http://repository.its.ac.id/id/eprint/91644>
- [11] S. H. Mohammed and S. Al-Augby, “LSA & LDA topic modeling classification: Comparison study on E-books,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, pp. 353–362, 2020, doi: 10.11591/ijeecs.v19.i1.pp353-362.

- [12] A. Athahahirah, "Pengaruh Ulasan Produk Terhadap Kepuasan Pengguna Aplikasi Female Daily," *INNOVATIVE: Journal Of Social Science Research*, vol. 3, pp. 7135–7147.
- [13] R. W. Pratiwi, S. F. H. Dairoh, D. I. Af'ida, Q. R. A., and A. G. F., "Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM)," *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol. 4, no. 1, pp. 40–46, 2021, doi: 10.20895/inista.v4i1.387.
- [14] U. T. Setijohatmo *et al.*, "Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik," *Prosiding Industrial Research Workshop and National Seminar*, vol. 11, no. 1, pp. 402–408, Sep. 2020, doi: <https://doi.org/10.35313/irwns.v11i1.2040>.
- [15] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis," *Front Artif Intell*, vol. 3, Jul. 2020, doi: 10.3389/frai.2020.00042.
- [16] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, Mar. 2003.
- [17] M. D. R Wahyudi, A. Fatwanto, U. Kiftiyani, and M. Galih Wonoseto, "Topic Modeling of Online Media News Titles during COVID-19 Emergency Response in Indonesia Using the Latent Dirichlet Allocation (LDA) Algorithm," *Telematika*, vol. 14, no. 2, pp. 101–111, Aug. 2021, doi: 10.35671/telematika.v14i2.1225.
- [18] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*. Association for Computing Machinery, Feb. 2015, pp. 399–408. doi: 10.1145/2684822.2685324.
- [19] A. Muhammin, S. R. S., and P. A. Atnanda, "Analisis Topic Modelling pada Ulasan Aplikasi Shopee di PlayStore Menggunakan Latent Dirichlet Allocation (LDA)," in *Prosiding Seminar Nasional Sains Data*, Nov. 2023, pp. 122–133. doi: 10.33005/senada.v3i1.91.
- [20] A. L. Lestari and A. Hananto, "How Do Firms Use Social Media: Topic Modeling of Twitter Brand Posts of Four Indonesian Skincare Brands," *ASEAN Marketing Journal*, vol. 15, no. 2, Dec. 2023, doi: 10.7454/amj.v15i2.1168.
- [21] A. F. A. Rohmaha, A. C. Fradani, and A. Indriani, "Pengaruh Electronic Word Of Mouth (E-WOM) Terhadap Keputusan Pembelian Pada Marketplace Tokopedia (Studi Pada Mahasiswa Pendidikan Ekonomi IKIP PGRI Bojonegoro)," *Jurnal Akuntansi Keuangan dan Bisnis*, vol. 1, no. 2, 2023, Accessed: Jun. 30, 2025. [Online]. Available: <https://jurnal.ittc.web.id/index.php/jakbs/article/view/74>
- [22] Erdawati, Endarwita, and R. Widiasari, "The Influence of Electronic Word of Mouth (E-WOM) and Product Quality on Skincare Purchase Decisions (Case Study of Generation Y in Lubuk Sikaping)," *Journal of Social and Economics Research*, vol. 5, no. 1, pp. 184–190, 2023, [Online]. Available: <https://idm.or.id/JSER/index>.