



Comparing LDA and BERTopic for Topic Modeling of Skin Complaints in Indonesian-Language Sunscreen Reviews on Shopee

Fahmi Ilmiah¹, Arif Faizin¹

¹Department of Informatic Engineering, Faculty of Engineering, Universitas Yudharta Pasuruan, Indonesi

*Corresponding Author: Fahmi Ilmiah

E-mail: ilmiahfahmi@gmail.com



Article Info

Article history:

Received 21 August 2026

Received in revised form 12 September 2026

Accepted 1 October 2026

Keywords:

Sunscreen

Topic Modeling

LDA

BERTopic

Tropical Climate

Abstract

Sunscreen use in Indonesia continues to grow alongside the high ultraviolet exposure typical of a tropical climate. However, various complaints such as a sticky texture, white cast, and the appearance of acne are still frequently found in user reviews. These reviews contain important information about users' experiences and skin problems, yet they are difficult to analyze manually because of their large volume and unstructured nature. This study therefore aims to identify skin-complaint patterns in sunscreen use through an analysis of user reviews on Shopee Indonesia. It applies a text mining approach using topic modeling methods, namely Latent Dirichlet Allocation (LDA) and BERTopic. Sunscreen product reviews were collected from the Shopee platform and then passed through preprocessing stages comprising cleaning, case folding, tokenizing, normalization, and stemming. The two methods were subsequently used to extract topics representing user complaints. Evaluation was carried out using the coherence score, topic diversity, and manual interpretation to assess the quality of the resulting topics. Based on the results, LDA produced nine topics with a coherence score of 0.35 and a topic diversity of 0.70, whereas BERTopic produced twelve topics with a coherence score of 0.50 and a topic diversity of 0.90. Based on the evaluation and an examination of dominant words and representative reviews, the topics produced by BERTopic showed better interpretability because they were more specific and easier to label, whereas LDA produced more general topics that still combined several complaints within a single topic. Accordingly,

Introduction

The skincare industry, particularly sunscreen, has grown rapidly in Indonesia as public awareness of self-care increases and access to digital information expands (Saparuddin et al., 2025; Syaputri & Herawati, 2025; Arigi et al., 20024). In this context, consumer reviews on e-commerce platforms have become an important consideration in purchase decisions (Nabila & Ratnasari, 2025) and, at the same time, generate large volumes of data for understanding consumer perceptions and needs (Nurcholis et al., 2026). The need for sunscreen is increasingly relevant given Indonesia's tropical climate, with high ultraviolet exposure throughout the year, which makes the skin prone to sunburn, premature aging, and UV-radiation damage (Fitraneti et al., 2024), as well as excessive oil production and acne caused by a hot and humid environment (Jung et al., 2019).

Nevertheless, choosing a sunscreen is relatively difficult because it must match skin type, sensitivity to active ingredients, and environmental conditions, while its usage rate remains relatively low (Fitraneti et al., 2024; Moloney et al., 2022). To examine these skin-complaint

patterns, this study analyzes sunscreen product reviews from the Indonesian local brands Wardah, Somethinc, Emina, and Azarine on Shopee the marketplace with the highest number of visits in Indonesia, at approximately 227.6 million visits per month in 2024 (Parasdita & Kurniasari, 2025). Wardah, for example, was recorded as the Sunscreen Aftersun brand with the largest market share, at 9.72% during the January–March 2026 period (Kompas, 2026), thereby generating a representative volume of reviews. Reviews on Shopee are generally short, unstructured, and written in informal language, which makes them both a challenge and an opportunity for the application of text mining (Wang et al., 2025; Allam et al., 2025; Vuković et al., 2025; Guenduez et al., 2026).

Analyzing large volumes of reviews can be carried out through text mining using topic modeling techniques, which can automatically extract the main themes from a collection of texts (Anjarsari et al., 2025; Huang & Jiang, 2026; Chu et al., 2022; Rüdiger et al., 2022). Latent Dirichlet Allocation (LDA) is a classical method that has proven effective in identifying topics from unstructured text and producing easily interpretable topics (Puspita et al., 2024), as well as providing information about user preferences and perceptions (Nabila & Ratnasari, 2025). BERTopic, as a modern embedding-based approach, is better able to preserve the semantic context between words than LDA (Anjarsari et al., 2025), making a comparison of the two important for determining the most optimal approach. Although topic modeling has been widely applied, several research gaps remain (Chen et al., 2023; Li et al., 2024; Beauté et al., 2026).

First, most studies use long, structured texts such as news articles or scientific documents, so the performance of topic modeling on short, informal, high-noise Indonesian text such as e-commerce reviews remains rarely examined (Nurcholis et al., 2026). Second, research in the skincare domain generally remains broad and has not specifically examined sunscreen products (Nabila & Ratnasari, 2025). Third, direct comparisons between a classical method (LDA) and a modern embedding-based method (BERTopic) on Indonesian-language consumer reviews are still limited (Nursyahrina et al., 2024). Fourth, model-quality evaluation in previous studies has generally focused only on topic coherence, without considering topic diversity and topic interpretability (Dieng et al., 2020). Fifth, studies linking review-analysis results to the tropical-climate context that affects skin conditions also remain limited (Putri et al., 2026; Putri et al., 2026; Demoze et al., 2024).

Based on these gaps, this study offers novelty through five main contributions: a focus on the specific domain of sunscreen products in Indonesia with the characteristics of tropical-climate skin problems; the use of an Indonesian-language consumer-review corpus from Shopee that represents real-world conditions (a real case); a comparative evaluation of LDA and BERTopic in the Indonesian-language skincare domain; topic interpretation based on skin-complaint patterns; and a more comprehensive model evaluation that considers topic coherence, topic diversity, and topic interpretability. Accordingly, this study aims to compare the performance of LDA and BERTopic in analyzing skin complaints in sunscreen product reviews on Shopee Indonesia and to identify the main topics representing skin problems in a tropical-climate region. Specifically, this study addresses two questions: (1) how do the topic-modeling results of LDA and BERTopic compare on an Indonesian-language sunscreen review corpus; and (2) which model performs better based on coherence score, topic diversity, and topic interpretability (Asnawi et al., 2024; Harahap et al., 2025; Mertayasa & Wulandharui, 2025).

These findings are expected to contribute academically to the development of text mining on Indonesian-language data and, practically, to help sunscreen manufacturers understand consumers' complaints and needs in a tropical climate.

Methods

Research Design

This study uses a quantitative, computational-comparative approach based on text mining to compare two topic modeling methods Latent Dirichlet Allocation (LDA) and BERTopic in identifying skin-complaint patterns in sunscreen product reviews on Shopee Indonesia. LDA is positioned as a probabilistic, Bag-of-Words baseline model, whereas BERTopic is a modern embedding-based model that better preserves semantic context. Comparing the two is relevant because e-commerce reviews are generally short, informal, unstructured, and noisy. All data processing was carried out using the Python programming language.

Data Collection and Unit of Analysis

The research object is consumer reviews (user-generated content) of sunscreen products on the Shopee Indonesia platform, treated as secondary data. Product selection used purposive sampling based on the following criteria: belonging to the sunscreen category; still actively sold at the time of data collection; having a sufficient number of reviews; containing comment text (not merely a rating); prioritizing Indonesian-language reviews; and being relevant to usage, suitability, texture, absorption, effects on the skin, or usage complaints. A total of 1,747 customer reviews were collected through a scraping and exporting process from five cosmetic brands, consisting of 1,007 reviews from Wardah, 296 from Azarine, 160 from Emina, 124 from Somethinc, and 160 from Skin Aqua, with review dates spanning on September 2020 until April 2026. Because Shopee reviews are dynamic and subject to access restrictions, the reviews were exported through a browser extension in CSV format and then processed in Google Colab. This process yielded 98 CSV files that were read without failure and 1,747 raw reviews. The unit of analysis is a single review, treated as a single text document.

Data Cleaning and Mapping

All CSV files were merged into a single dataset with standardized column names (comment, rating, date, product name, brand, and product URL). The data-cleaning stage included removing duplicates, unnecessary special characters, numbers, and symbols, as well as empty or incomplete data. The dataset was then mapped into several groups specific skin complaints, negative non-skin comments, logistics/seller comments (seller noise), and general/neutral comments so that only reviews relevant to skin complaints entered the modeling stage. Complaint mapping was based on a list of keywords per category, for example acne/breakout, irritation/itch/redness, oily skin, white cast, sticky/slow-absorbing texture, dry/dull skin, incompatibility, and heat/sunburn.

To improve classification accuracy, negation handling based on the concept of negation scope detection was applied (Councill et al., 2010). Negation variants such as *ga*, *gak*, and *nggak* were normalized to *tidak* (not); the system then checked whether a complaint keyword fell within the negation scope of the same clause, and if so, that keyword was not counted as a complaint. A clause-boundary rule was applied so that the effect of negation did not extend to the following clause for example, in “*tidak lengket, tetapi bikin jerawat*” (not sticky, but causes acne), the word *jerawat* (acne) is still recognized as a complaint. Conversely, phrases such as *tidak cocok* (not suitable) and *tidak meresap* (does not absorb) are still classified as complaints because, taken as a whole, they represent a negative user experience. Negation handling was performed before preprocessing so that negation words were not removed during stopwords removal

Text Preprocessing

Preprocessing for LDA comprised case folding, character cleaning, limited normalization, tokenizing, stopwords removal, and stemming using Sastrawi, resulting in a tokenized corpus.

For BERTopic, preprocessing was lighter, without aggressive stemming, so that the semantic context was preserved (Aban & Ratnasari, 2026).

Topic Modeling

In LDA, the preprocessing results were used to build a dictionary and corpus, the text was represented using Bag of Words, and the optimal number of topics was determined based on the coherence value. The model produces a topic distribution for each document and a word distribution for each topic. BERTopic was run through four stages: representing the text with transformer-based embeddings, reducing dimensionality using UMAP, clustering using HDBSCAN, and representing topics using c-TF-IDF to obtain the most distinctive words for each cluster (Grootendorst, 2020). The embedding model used was *paraphrase-multilingual-MiniLM-L12-v2* because it supports multilingual text, including Indonesian. The similarity between embeddings was measured using cosine similarity, $\cos \theta = (w \cdot x) / (|w| \cdot |x|)$, which approaches 1 as two vectors become more similar. The distinctiveness weight of a term was computed using c-TF-IDF, $c\text{-TF-IDF}(t,c) = \text{tf}(t,c) \times (1 + A/f_t)$. [Model configuration: specify the LDA hyperparameters (α , β) and the UMAP/HDBSCAN parameters such as `n_neighbors` and `min_cluster_size`, or state that default values were used.]

Topic Evaluation and Labeling

The performance of LDA and BERTopic was evaluated across three aspects. First, the coherence score (C_v , based on NPMI with a sliding-window size of 110) measures the semantic relatedness between words within a topic; a higher value indicates a more coherent topic. Second, topic diversity measures the variety of representative words across topics, namely the proportion of unique terms in the combined top words of all topics; a value approaching 1 indicates little term repetition across topics. Third, interpretability was assessed qualitatively through manual validation: each topic was labeled based on the combination of dominant words (top words), word weights, and representative reviews, and then verified to ensure it reflected the intended skin-complaint meaning in context. These three aspects formed the basis for determining the more effective method for extracting skin-complaint topics.

Research Flow

This study began with a scraping process to collect sunscreen product reviews from the Shopee Indonesia platform. The collected data were stored in several CSV files according to the data-retrieval process. All CSV files were then merged into a single, complete dataset. The next stage was data cleaning, which aimed to improve data quality before analysis; this process included removing duplicate data, handling missing values, and cleaning unnecessary characters or information so that the dataset became more consistent and ready for use. After cleaning, data mapping was performed to group the reviews by research category; at this stage, reviews were separated into comments related to skin complaints or usage effects and comments unrelated to skin complaints.

The next stage was text preprocessing, that is, converting the text data into a more structured form before modeling. Preprocessing included case folding, cleaning of unnecessary characters, word normalization, stopword removal, tokenization, and stemming. The preprocessed data were then analyzed using two topic modeling methods, namely Latent Dirichlet Allocation (LDA) and BERTopic. Both methods were applied to the same dataset to produce a set of topics representing the discussion patterns in sunscreen users' reviews. The final stage was topic evaluation and labeling. The results of each method were evaluated using coherence score, topic diversity, and topic interpretability through manual validation. Each topic was then labeled based on the representative words (top words) produced, yielding the main theme of each topic. These evaluation results were used as the basis for comparing the

performance of LDA and BERTopic in extracting topics from sunscreen product reviews on Shopee Indonesia

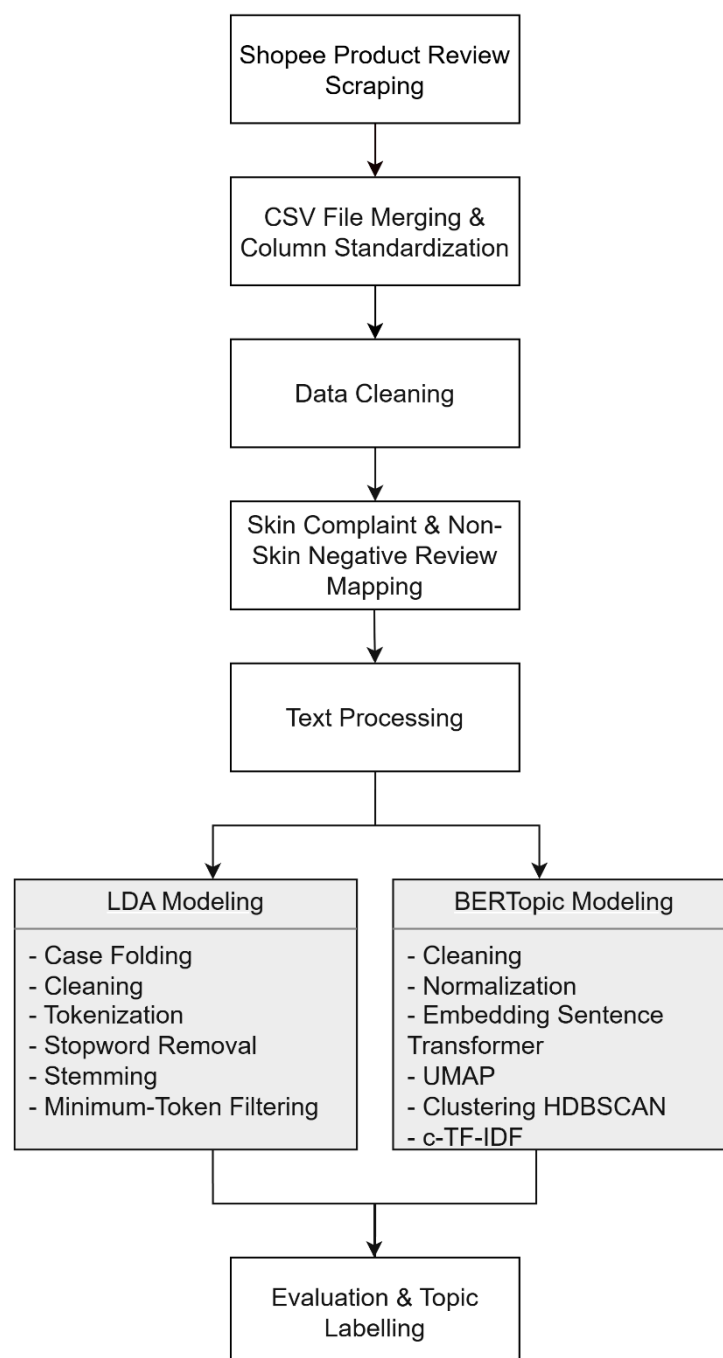


Figure 1. Research Flow

Results and Discussion

Data Overview and Comment Mapping

The raw data scraped from Shopee (using the Shopee Reviews Extractor) amounted to 1,747 reviews. After merging the datasets, standardizing the columns, and performing the initial cleaning, 1,523 out of 1,747 raw scraped records were found to contain non-empty comments, while 224 records contained missing comments. After removing empty comments and duplicate entries, 843 unique comments remained. Subsequently, a relevance filter was applied to identify comments related to skin conditions and concerns, using keywords such as acne, breakouts, oily skin, and other relevant terms. This filtering process resulted in a final

987

dataset of 619 relevant reviews for further analysis. These reviews were mapped into three groups 385 specific skin-complaint comments, 147 negative non-skin comments, and 87 general comments (Table 1). The 385 specific skin-complaint comments were selected for further analysis because they directly addressed consumers' experiences, concerns, and complaints related to skin conditions, making them the most relevant subset for identifying underlying topics and patterns in skin-related product reviews. In addition, the number of skin-complaint comments (385) exceeded the minimum threshold of 150 documents, providing a sufficient volume of text for reliable topic modeling. Therefore, this group was used as the final dataset for LDA and BERTopic modeling. Documents containing fewer than three tokens were discarded to ensure that the remaining documents provided sufficient textual information for meaningful topic representation.

Table 1. Distribution of Comment Types

Data Component	Count	Description
Specific skin complaints	385	Subset prioritized for LDA and BERTopic
Negative non-skin comments	147	Negative comments about quality, price, seller, logistics, or non-skin purchase experience
General comments	87	Comments set aside so as not to interfere with the skin-complaint analysis

Source: Processed by the authors, 2026

Distribution of Skin-Complaint Categories

The mapping of skin-complaint categories produced eight categories, as shown in Figure 2. The mapping shows that the most dominant complaint category is a sticky, heavy, or slow-absorbing texture, with 330 occurrences, or 69.77% of the total skin-complaint category occurrences. The dominance of this category indicates that usage comfort is a primary concern for sunscreen users. The hot and humid tropical climate is thought to influence users' perceptions of product texture, so that sunscreens that feel sticky or heavy, or that take a long time to absorb, are more frequently criticized. The next category is oily skin, with 47 occurrences (9.94%). This complaint is also related to the tropical climate, which can increase oil production on the skin, so that certain sunscreens may produce a more oily sensation on the face. Next, irritation, stinging, itching, or redness was found in 28 occurrences (5.92%), while acne, small bumps, breakouts, or comedones were recorded in 27 occurrences (5.71%). These two categories indicate that some users experienced reactions related to skin condition after using sunscreen.

In addition, the dry or dull skin category was found in 18 occurrences (3.81%), followed by product incompatibility in 11 occurrences (2.33%) and white cast/greyish/cakey appearance, also 11 occurrences (2.33%). The category with the fewest occurrences was stinging eyes, with 1 occurrence (0.21%). Overall, the mapping shows that users' complaints about sunscreen relate not only to skin problems such as acne, irritation, or dryness, but are also dominated by the product-usage experience, particularly a texture that feels sticky, heavy, or slow to absorb, as well as an oily effect after application. This finding indicates that usage comfort is one of the important factors influencing users' assessments of sunscreen products, especially under tropical-climate conditions

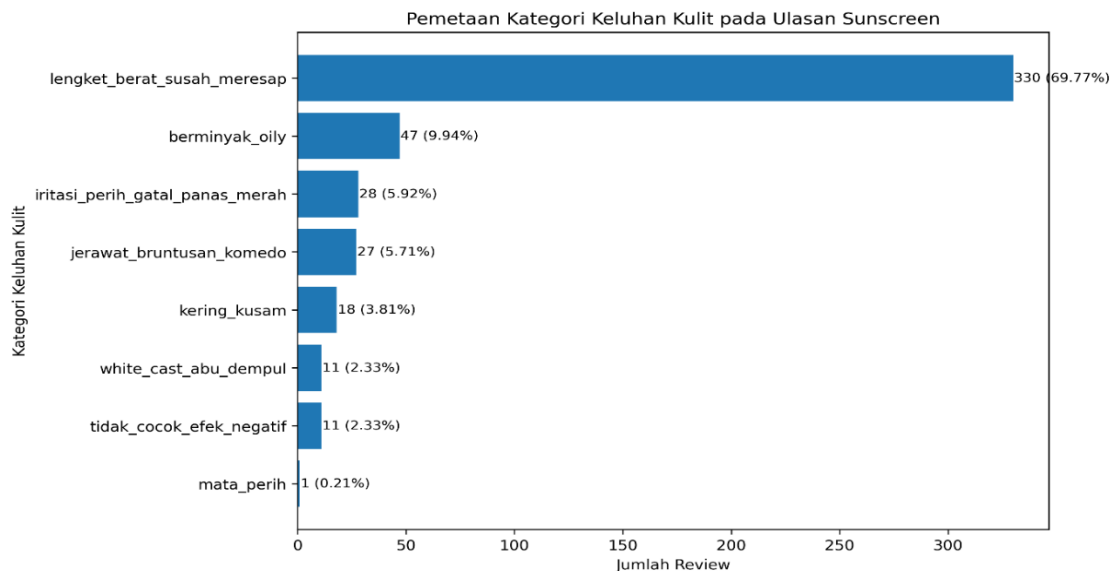


Figure 2. Distribution of Skin-Complaint Categories

LDA Modeling Results

The best LDA model produced 9 topics with a coherence score of 0.35 and a topic diversity of 0.70 (Table 2). During the LDA stage, the model was tested with several numbers of topics, each evaluated using the coherence score and topic diversity metrics; after the dictionary and corpus were built, the 9-topic model obtained the best coherence score of 0.35 and a topic diversity of 0.70.

Table 2. LDA Modeling Results

Topic	Top Words (with weights)	Label
0	aroma:0.0336; fast:0.0240; face:0.0222; smell:0.0162; courier:0.0155; here:0.0151; oil:0.0146; wardah:0.0142; heavy:0.0126; pleasant:0.0125; effect:0.0122; somewhat:0.0122	Oily-skin complaint
1	rays:0.0402; protect:0.0358; gel:0.0281; sunlight:0.0226; effect:0.0181; heavy:0.0172; azarine:0.0144; face:0.0126; want:0.0117; whitecast:0.0114; spf:0.0102; directly:0.0100	Sticky/heavy/slow-absorbing texture complaint
2	application:0.0236; liquid:0.0203; without:0.0159; sell:0.0150; moist:0.0147; response:0.0141; face:0.0140; heavy:0.0121; suitable:0.0120; open:0.0118; fragrant:0.0118; dry:0.0118	Sticky/heavy/slow-absorbing texture complaint
3	quality:0.0455; satisfied:0.0328; reach:0.0259; suitable:0.0219; sticky:0.0185; cream:0.0181; regular-customer:0.0180; service:0.0166; time:0.0165; usual:0.0149; large:0.0143; gel:0.0131	Sticky/heavy/slow-absorbing texture complaint
4	oil:0.0183; gel:0.0180; absorb:0.0179; usual:0.0177; absorption:0.0165; order:0.0164; packaging:0.0157; heavy:0.0155; aroma:0.0141; matter:0.0124; will:0.0121; blend:0.0121	Oily-skin complaint
5	gold:0.0321; fast:0.0234; oil:0.0229; gel:0.0185; liquid:0.0183; very:0.0125; good:0.0114;	Oily-skin complaint

	really:0.0114; content:0.0107; satisfied:0.0105; already:0.0099; open:0.0098	
6	oil:0.0283; liquid:0.0231; face:0.0197; seal:0.0166; enough:0.0160; indeed:0.0146; leaking:0.0138; always:0.0126; protect:0.0114; packaging:0.0114; difficult:0.0108; if:0.0105	Oily-skin complaint
7	order:0.0357; liquid:0.0327; suitable:0.0300; heavy:0.0214; thank-you:0.0195; packaging:0.0152; thankfully:0.0145; always:0.0126; complete:0.0124; disappointed:0.0124; many:0.0121; open:0.0108	Sticky/heavy/slow-absorbing texture complaint
8	acne:0.0499; pimples:0.0420; prone:0.0291; wardah:0.0279; spf:0.0255; oil:0.0202; calming:0.0194; skin:0.0192; effect:0.0171; smooth:0.0120; application:0.0115; pores:0.0114	Acne/breakout/comedone complaint

Source: Processed by the authors, 2026

Based on the table, Topic 0 is dominated by the words *aroma*, *muka* (face), *minyak* (oil), *wardah*, *berat* (heavy), and *efek* (effect). The presence of *minyak* and *berat* indicates that this topic discusses oily-skin complaints after sunscreen use, while *aroma* and *bau* (smell) appear as supporting information about the user's experience with the product.

Topic 1 contains the words *sinar* (ray), *lindung* (protect), *gel*, *matahari* (sun), *efek*, *berat*, and *whitecast*. Although several words relate to the sunscreen's function, the presence of *berat* and *whitecast* indicates that this topic mostly discusses complaints about a product texture that feels heavy and leaves a white cast after application.

Topic 2 is dominated by the words *cair* (runny), *lembab* (moist), *muka*, *berat*, and *kering* (dry). This combination describes users' experience of the sunscreen's texture, particularly application comfort, the ability to moisturize the skin, and complaints about a product that feels heavy or uncomfortable to use.

Topic 3 contains the words *lengket* (sticky), *cream*, *gel*, *kualitas* (quality), and *sesuai* (suitable). The prominence of *lengket* indicates that this topic focuses on complaints about a sunscreen texture that feels sticky after use, while the other words relate more to users' assessments of the product.

Topic 4 is dominated by the words *minyak*, *gel*, *nyerap* (absorb), *resap* (absorb), and *berat*. This topic describes oily-skin complaints related to the sunscreen's ability to absorb into the skin. Users complained about products that feel heavy, are difficult to absorb, and leave an oily impression after use.

Topic 5 contains the words *minyak*, *gel*, *cair*, and *cepat* (fast). Although some words are general, the presence of *minyak* indicates that this topic still relates to oily-skin complaints after sunscreen use.

Topic 6 is dominated by the words *minyak*, *cair*, *muka*, *lindung*, and *susah* (difficult). This topic again reflects discussion of skin becoming oilier after using sunscreen. Other words, such as *kemas* (packaging) and *segel* (seal), appear as supporting information about the user's product experience.

Topic 7 contains the words *cair*, *berat*, *kemas*, and *kecewa* (disappointed). The presence of *berat* as one of the dominant words indicates that this topic relates to complaints about a sunscreen texture that feels heavy when used, although some general words not directly related to skin complaints also appear.

Topic 8 is dominated by the words acne, jerawat (acne), prone, calming, skin, and minyak. This topic clearly describes complaints about acne, small bumps, or acne-prone skin after sunscreen use. The words calming and skin indicate that some users also discussed the product's effect on acne-prone skin conditions.

Based on the manual interpretation of the topic modeling results, the first author assigned labels to the identified topics according to their representative words and contextual meaning. The assigned labels were subsequently validated by one of the authors' academic supervisors to assess their consistency and relevance to the review content. The LDA model produced nine topics dominated by three main complaint groups: sticky, heavy, or slow-absorbing texture; oily skin; and acne or small bumps. Most topics still show general words such as gel, cair, muka, and kemas appearing in more than one topic, so the boundaries between topics are not yet fully separated. This indicates that LDA can group the main themes in sunscreen reviews, but some topics still overlap, so topic interpretation tends to be more general than a specific separation of complaints

BERTopic Modeling Results

The best BERTopic model produced 12 topics with a coherence score of 0.50 and a topic diversity of 0.90 (Table 3). The model was built using an embedding technique to represent the semantic meaning of the text. Documents that had undergone light preprocessing were converted into embeddings using the paraphrase-multilingual-MiniLM-L12-v2 model. Dimensionality reduction was then performed using UMAP, and the documents were clustered using HDBSCAN. After the clusters were formed, CountVectorizer was used to build a word-frequency representation, which was then processed using c-TF-IDF to obtain the representative words for each topic. The BERTopic results were stored as topic information, a list of top words with their weights, initial suggested labels, and representative reviews. Error handling was also provided so that the cause of runtime failures could be identified, especially when Google Colab resources were insufficient.

Table 3. BERTopic Modeling Results

Topic	Top Words (with weights)	Label
0	ray:0.0563; protect:0.0457; sun:0.0414; sunlight:0.0414; protect from rays:0.0380; sensitive:0.0363; oily:0.0348; effect:0.0346; wardah:0.0343; really:0.0318	Oily-skin complaint
1	aroma:0.1086; cream:0.0693; not:0.0631; liquid:0.0542; sticky:0.0407; fragrant:0.0384; absorb:0.0343; pleasant:0.0338; content:0.0301; come out:0.0301	Sticky/heavy/slow-absorbing texture complaint
2	authenticity:0.0549; quality:0.0522; arrive:0.0506; small bumps:0.0495; acne-prone:0.0493; opened:0.0451; not:0.0445; small:0.0435; different:0.0379; smooth:0.0379	Acne/breakout/comedone complaint
3	heavy:0.1935; sticky:0.0747; azarine:0.0572; period:0.0498; heavy period:0.0451; valid:0.0451; whitecase:0.0451; validity period:0.0451; heavy 2028:0.0451; somewhat:0.0447	Sticky/heavy/slow-absorbing texture complaint
4	not:0.0693; dry:0.0563; acne-prone:0.0478; dull:0.0450; authenticity:0.0407; absorb:0.0404; honest:0.0354; pleasant:0.0353; oily:0.0344; face:0.0333	Oily-skin complaint

5	basically:0.0470; wardah:0.0445; as good as:0.0436; whitecast:0.0416; get:0.0399; fast:0.0381; hard to blend:0.0373; ok:0.0371; other:0.0352; shopee:0.0336	White cast/greyish/cakey complaint
6	satisfied:0.1015; discount:0.0648; affordable:0.0629; service:0.0567; high-quality:0.0531; slow absorption:0.0517; time:0.0508; suitable:0.0495; customer:0.0443; difficult to absorb:0.0429	Sticky/heavy/slow-absorbing texture complaint
7	acne:0.1198; spf:0.0585; calming:0.0552; skin:0.0502; acne appears:0.0484; oily:0.0473; 35:0.0417; help:0.0417; spf 35:0.0417; wardah:0.0403	Acne/breakout/comedone complaint
8	ok:0.1106; ok ok:0.0989; packaging:0.0764; speed:0.0688; low absorption:0.0688; difficult to absorb:0.0616; quality:0.0611; response:0.0568; shipped:0.0555; product:0.0555	Sticky/heavy/slow-absorbing texture complaint
9	cardboard box:0.1007; packaging:0.0845; leaking:0.0717; additional:0.0683; even though:0.0525; heavy texture:0.0525; liquid:0.0509; order:0.0485; wrapped:0.0483; home:0.0456	Sticky/heavy/slow-absorbing texture complaint
10	disappointed:0.0874; box:0.0776; indeed:0.0718; causes acne:0.0718; small bumps:0.0626; hopefully:0.0576; not:0.0550; air:0.0543; told to:0.0543; use:0.0517	Acne/breakout/comedone complaint
11	clear:0.0817; why:0.0613; where:0.0572; ok:0.0570; unsuitable:0.0554; shipped:0.0515; even though:0.0474; an:0.0474; ok ok:0.0459; with:0.0447; finished:0.0422	Product-incompatibility complaint

Source: Processed by the authors, 2026

Based on the BERTopic modeling, Topic 0 is dominated by the words sinar, melindungi (protect), matahari, sensitif (sensitive), and berminyak (oily). These words indicate that users' discussion relates to the sunscreen's function of protecting the skin from sunlight, but some users also complained about skin becoming oilier after use. Based on manual validation, this topic is interpreted as an oily-skin complaint.

Topic 1 is dominated by the words aroma, cream, cair, lengket, menyerap (absorb), and wangi (fragrant). This combination indicates that users discussed the sunscreen's texture characteristics extensively, particularly stickiness and the product's absorption into the skin. Therefore, this topic is labeled a sticky, heavy, or slow-absorbing texture complaint.

In Topic 2, dominant words such as beruntusan (small bumps), berjerawat (acne-prone), keaslian (authenticity), kualitas, and halus (smooth) indicate that some users associated sunscreen use with the appearance of acne or small bumps. Based on manual validation, this topic represents an acne, small-bump, or comedone complaint.

Topic 3 is dominated by the words berat, lengket, azarine, and whitecase. These words indicate that the discussion focuses on a sunscreen texture that feels heavy, sticky, and less comfortable to use. Based on manual interpretation, this topic is categorized as a sticky, heavy, or slow-absorbing texture complaint.

Topic 4 contains the words *kering*, *berjerawat*, *kusam* (dull), *menyerap*, *berminyak*, and *muka*. These words indicate a discussion of changes in skin condition after sunscreen use, particularly skin that feels oilier, although some other complaints such as dry or dull skin also appear. Based on manual validation, this topic is interpreted as an oily-skin complaint.

In Topic 5, words such as *whitecast*, *susah blend* (hard to blend), *wardah*, and *cepet* indicate that users discussed the white-cast effect and the difficulty of blending the product on the face. Therefore, this topic is labeled a white cast, greyish, or cakey-effect complaint.

Topic 6 is dominated by the words *lama meresap* (slow to absorb), *susah nyerap* (hard to absorb), *pelayanan* (service), *diskon* (discount), and *terjangkau* (affordable). Although some words relate to service and price, the phrases *lama meresap* and *susah nyerap* are the main indicators, so this topic represents a sticky, heavy, or slow-absorbing texture complaint.

In Topic 7, dominant words such as *acne*, *timbul jerawat* (acne appearing), *calming*, *skin*, and *berminyak* indicate that users discussed the appearance of acne after sunscreen use, particularly on acne-prone skin. Based on manual validation, this topic is categorized as an acne, small-bump, or comedone complaint.

Topic 8 contains the words *daya serap rendah* (low absorption), *susah serap* (hard to absorb), *pengemasan* (packaging), *kecepatan* (speed), and *kualitas*. Although some words relate to packaging and shipping, the main focus remains on complaints about the sunscreen's difficulty absorbing into the skin. Therefore, this topic is interpreted as a sticky, heavy, or slow-absorbing texture complaint.

In Topic 9, dominant words such as *tekstur berat* (heavy texture), *cair*, *kemasan* (packaging), *bocor* (leaking), and *kardus* (cardboard) indicate that users discussed the product-usage experience together with the condition of the packaging received. Based on manual validation, this topic is still categorized as a sticky, heavy, or slow-absorbing texture complaint, because *tekstur berat* is the most relevant to the research focus.

Topic 10 is dominated by the words *bikin jerawat* (causes acne), *beruntusan*, *kecewa*, and *kotak* (box). These words indicate that some users experienced the appearance of acne or small bumps after using sunscreen, accompanied by comments about the purchase experience. Based on manual interpretation, this topic represents an acne, small-bump, or comedone complaint.

Topic 11 contains the words *tidakcocok* (not suitable), *jelas* (clear), *mana*, *ok*, and *dikirim* (shipped). The presence of *tidakcocok* is the main indicator that users complained about the product's incompatibility with their skin condition. Therefore, this topic is labeled a product-incompatibility complaint.

Based on the manual interpretation, BERTopic produced 12 topics that describe the variety of sunscreen users' complaints in greater detail. The most dominant topics relate to sunscreen texture, such as stickiness, heaviness, and slow absorption. In addition, BERTopic successfully separated more specific topics, including oily skin, acne or small bumps, white cast, and product incompatibility. Although some words related to non-skin aspects such as packaging, service, or shipping still appear, they serve as supporting context and do not dominate the topic interpretation. These results show that BERTopic can produce a more specific topic representation, making labeling and interpretation easier than approaches that yield more general topics.

Comparison of LDA and BERTopic

The comparison of the two methods is summarized in Table 4. LDA produced 9 topics (coherence 0.35; diversity 0.70), whereas BERTopic produced 12 topics (coherence 0.50; diversity 0.90). Of the total of 21 topics produced by the two methods, LDA accounted for

approximately 42.86% and BERTopic 57.14%. BERTopic was superior across all three aspects number of topics, coherence, and diversity and produced topics that were easier to interpret, although still requiring manual validation.

Table 4. Comparison of LDA and BERTopic

Model	Number of Topics	Coherence Score	Topic Diversity	Manual Interpretation
LDA	9	0.35	0.70	Topics tend to be general; several complaint types remain combined within a single topic
BERTopic	12	0.50	0.90	Topics are more specific; each topic is easier to label and represents a single complaint type

Source: Processed by the authors, 2026

Based on the modeling results, LDA produced 9 topics, whereas BERTopic produced 12 topics. When compared by the number of topics produced, LDA accounted for approximately 42.86% of the total of 21 topics, while BERTopic accounted for approximately 57.14%. This difference indicates that BERTopic can provide a more detailed division of topics than LDA. To evaluate the quality of the generated topics and select the optimal model, two complementary metrics were used: coherence score and topic diversity. Coherence score measures the semantic relatedness among words within a topic, indicating how coherent, interpretable, and meaningful the resulting topics are. Topic diversity measures the proportion of unique words across all topics, reflecting the extent to which the identified topics are distinct and reducing the occurrence of overlapping words across topics (Dieng et al., 2020; Weston et al., 2022; Grassia et al., 2024). Therefore, both metrics were considered simultaneously to identify a model that produces coherent and interpretable topics while maintaining sufficient diversity between topics. In terms of evaluation, LDA obtained a coherence score of 0.35 and a topic diversity score of 0.70, whereas BERTopic obtained a coherence score of 0.50 and a topic diversity score of 0.90. The higher scores obtained by BERTopic on both metrics indicate that its topics were more semantically coherent and more diverse than those generated by LDA. Based on these evaluation results, BERTopic was selected as the optimal model for further topic analysis.

Based on the manual interpretation conducted by the first author, the dominant words and representative reviews were examined to assess the meaning and interpretability of the topics generated by LDA. This manual interpretation was necessary because coherence score and topic diversity provide quantitative measures of topic quality but do not fully indicate whether the resulting topics represent clear and specific complaint themes. The manual interpretation showed that the topics produced by LDA remained relatively general, with several types of skin complaints being combined within a single topic. This finding suggests that, despite its quantitative evaluation, LDA had limitations in distinguishing between specific complaint categories. Therefore, manual interpretation was used to complement the quantitative evaluation by assessing whether the generated topics were sufficiently clear, specific, and meaningful in the context of the reviews. In contrast, BERTopic produced more specific topics that were easier to label and able to represent a single complaint type. Accordingly, based on coherence score, topic diversity, and manual interpretation, BERTopic demonstrated better performance than LDA in analyzing the corpus of Indonesian-language sunscreen product reviews.

The mapping shows that the most dominant complaints relate to a sunscreen texture that is sticky, heavy, or slow to absorb. This finding is important in the tropical-climate context because high heat and humidity can make users more sensitive to sticky and heavy sensations

on the face. Sunscreen that does not absorb quickly may also cause discomfort during daily activities. Academically, this study shows that topic-modeling analysis of sunscreen reviews must take into account the characteristics of short text, noise, and domain context. Mapping comments before modeling is important so that the resulting topics are not mixed with logistics or seller comments. This strengthens the study's contribution to applying text mining to Indonesian-language e-commerce reviews. Practically, the results can provide input for sunscreen manufacturers: the dominant complaints about texture, absorption, oiliness, white cast, and skin incompatibility indicate that consumers need products that are lightweight, quick-absorbing, not too oily, minimal in white cast, and suitable for skin conditions in a tropical climate

Conclusion

Based on the data processing and modeling, it can be concluded that sunscreen product reviews on Shopee Indonesia contain a variety of information about skin complaints and usage experiences. After mapping, 385 specific skin-complaint comments were found out of 619 main reviews. Based on the mapping of skin-complaint categories, the most dominant complaint is a sticky, heavy, or slow-absorbing texture. In addition, oily-skin, irritation/stinging/itching/redness, acne/small-bump/breakout, dry or dull skin, product-incompatibility, white cast/greyish/cakey, and stinging-eye complaints were also found, with fewer occurrences.

The comparison of the Latent Dirichlet Allocation (LDA) and BERTopic methods shows differences in both the number and the characteristics of the topics produced. LDA produced nine topics with a coherence score of 0.35 and a topic diversity of 0.70. Based on manual interpretation, the resulting topics tended to be general, so that several complaint types were still combined within a single topic. Meanwhile, BERTopic produced twelve topics with a coherence score of 0.50 and a topic diversity of 0.90. The manual interpretation shows that the topics produced by BERTopic are more specific, easier to label, and better able to represent a single complaint type.

Based on the evaluation using coherence score, topic diversity, and topic interpretability, BERTopic demonstrated better performance than LDA in extracting topics from the corpus of sunscreen product reviews on Shopee Indonesia. The higher coherence score and topic diversity, together with more specific and easily interpretable topics, indicate that BERTopic is more effective in identifying skin-complaint patterns in Indonesian-language review data. Nevertheless, the difference in the number of topics produced and the limited amount of research data remain factors to consider when interpreting the results.

Suggestion

Future research can expand the number of sunscreen products and reviews so that the resulting topics are more representative. In addition, future work can combine topic modeling with sentiment analysis to determine whether each topic tends toward positive, neutral, or negative sentiment. Future research can also apply stronger validation using two independent raters to label topics and compute inter-rater agreement. This would strengthen the interpretability of the topic-modeling results, especially when the resulting topics still contain general words or overlap between categories. Furthermore, research can compare different embedding models in BERTopic, for example a multilingual sentence transformer and an Indonesian-specific model, so that the quality of topics on Indonesian-language review text can be evaluated more comprehensively.

References

Aban, N. A., & Ratnasari, C. I. (2026). Pemodelan Topik Cuitan Tentang Danantara Menggunakan Bertopic Teroptimasi Umap Dan Hdbscan. *Rabit : Jurnal Teknologi*

- Allam, H., Makubvure, L., Gyamfi, B., Graham, K. N., & Akinwolere, K. (2025). Text classification: How machine learning is revolutionizing text categorization. *Information*, 16(2), 130. <https://doi.org/10.3390/info16020130>
- Anjarsari, D. S., Imaningsih, A. N., Arviani, I., & Prasetyo, A. Y. (2025). Hubungan Penggunaan Face Mist Berbahan Dasar Alami Terhadap Kesehatan Kulit Wajah pada Cuaca Ekstrem Di kalangan Masyarakat Kabupaten Jombang. *Journal Of Social Science Research*, 5, 1090–1102.
- Arigi, J. P., Hidayat, Z., Putra, D. O., & Gulo, A. (2024). Analysis of metrosexual consumer lifestyle in urban areas: A phenomenological study of male-targeted beauty consumers on social media. *Humanities, Arts and Social Sciences Studies*, 768-780. <https://doi.org/10.69598/hasss.24.3.264857>
- Asnawi, M. F., Hanafi, M., Kurniawan, N. F., Suwondo, A., Nasrullah, A., & Setyawan, C. (2024, November). Topic modelling analysis on indonesian news using bert topic model. In *2024 6th International Conference on Cybernetics and Intelligent System (ICORIS)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ICORIS63540.2024.10903779>
- Beauté, R., Schwartzman, D. J., Dumas, G., Crook, J., Macpherson, F., Barrett, A. B., & Seth, A. K. (2026). Mapping of Subjective Accounts into Interpreted Clusters (MOSAIC): topic modelling and LLM applied to stroboscopic phenomenology. *Neuroscience of Consciousness*, 2026(1), niag008. <https://doi.org/10.1093/nc/niag008>
- Chen, Y., Peng, Z., Kim, S. H., & Choi, C. W. (2023). What we can do and cannot do with topic modeling: A systematic review. *Communication Methods and Measures*, 17(2), 111-130. <https://doi.org/10.1080/19312458.2023.2167965>
- Chu, K. E., Keikhosrokiani, P., & Asl, M. P. (2022). A topic modeling and sentiment analysis model for detection and visualization of themes in literary texts. *Pertanika Journal of Science & Technology*, 30(4), 2535-2561.
- Councill, I. G., Mcdonald, R., & Velikovich, L. (2010). What ' s Great and What ' s Not : Learning to Classify the Scope of Negation for Improved Sentiment Analysis. July, 51–59.
- Demoze, L., Gubena, F., Akalewold, E., Brhan, H., Adane, K. C., Kifle, T., ... & Yitageasu, G. (2024). Burden and determinants of scabies in Ethiopian school age children: A systematic review and meta-analysis with public health implications. *PloS one*, 19(12), e0314882. <https://doi.org/10.1371/journal.pone.0314882>
- Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8, 439–453. https://doi.org/10.1162/tacl_a_00325
- Fitraneti, E., Rizal, Y., Riska Nafiah, S., Primawati, I., & Ayu Hamama, D. (2024). Pengaruh Paparan Sinar Ultraviolet terhadap Kesehatan Kulit dan Upaya Pencegahannya : Tinjauan Literatur. *Scientific Journal*, 3(3), 185–194. <https://doi.org/10.56260/sciena.v3i3.147>
- Grassia, M. G., Marino, M., Mazza, R., Misuraca, M., Zavarrone, E., & Friel, M. (2024). Regional Competitiveness: A Structural-Based Topic Analysis on Recent Literature: MG Grassia et al. *Social Indicators Research*, 173(1), 83-108. <https://doi.org/10.1007/s11205-022-02951-4>

- Grootendorst, M. (2020). BERTopic: Neural topic modeling with a class-based TF-IDF procedure.
- Guenduez, A. A., Fuchs, S., Mergel, I., Mettler, T., & Frowein, S. (2026). From patterns to meaning: a mixed-methods framework that integrates computational and qualitative text analysis. *Quality & quantity*, 60(1), 2713-2738. <https://doi.org/10.1007/s11135-025-02357-7>
- Harahap, D. K., Tania, K. D., & Sevtiyuni, P. E. (2025). Topic Mining-Based Knowledge Discovery of User Health Information Needs. *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 7(4), 641-653. <https://doi.org/10.35882/ijeemi.v7i4.270>
- Huang, Z., & Jiang, Z. (2026). Text mining of syntactic complexity in L2 writing: An LDA topic modeling approach. *International Review of Applied Linguistics in Language Teaching*, 64(1), 523-548.
- Jung, I. S., Yun, S. J., & Lee, J. B. (2019). *The Difference in Sebum Secretion Affecting Development of Acne*. *Annals of Dermatology*, 31(4), 426-433.
- Li, D., Wu, K., & Lei, V. L. (2024). Applying topic modeling to literary analysis: A review. *Digital Studies in Language and Literature*, 1(1-2), 113-141.
- Mertayasa, I. K. T., & Wulandhari, L. A. (2025, November). Topic Modeling of Indonesian-Language Travel Reviews on Things to Do in Bali Using BERTopic and IndoBERT. In *2025 13th International Conference on Awareness Science and Technology (iCAST)* (pp. 1-7). IEEE. <https://doi.org/10.1109/iCAST68191.2025.11300049>
- Moloney, F. J., Collins, S., & Murphy, G. M. (2002). Sunscreens: safety, efficacy and appropriate use. *American journal of clinical dermatology*, 3(3), 185-191. <https://doi.org/10.2165/00128071-200203030-00005>
- Nabila, N., & Ratnasari, C. I. (2025). Topic Modeling of Skincare Comments from Female Daily. *Journal of Applied Informatics and Computing*, 9(4), 1394-1405. <https://doi.org/10.30871/jaic.v9i4.9625>
- Nurcholis, J. D., Hadibrata, M., Arzakayah, M. M., Nasution, F. B., & Yuliandari, P. (2026). Analisis Big Data pada Perdagangan Digital: Studi Komparatif E-Commerce, S-Commerce, dan Q-Commerce Berbasis BERTopic. *INOMATEC: Jurnal Inovasi dan Kajian Multidisipliner Kontemporer*, 1(3), 417-431.
- Nursyahrina, Defit, S., & Sovia, R. (2024). Metode BERTopic dan LDA untuk Analisis Tren Penelitian Bidang Ilmu Komputer. *Jurnal KomtekInfo*, 11(4), 332-341. <https://doi.org/10.35134/komtekinfo.v11i4.580>
- Parasdita, A. C., & Kurniasari, F. (2025). The Influence of Promotion , Online Customer Review , and Online Customer Rating on Purchasing Decisions in the Shopee Application Among Students in Pontianak City. 6(2), 1-9.
- Puspita, E., Shiddieq, D. F., & Roji, F. F. (2024). Pemodelan Topik pada Media Berita Online Menggunakan Latent Dirichlet Allocation (Studi Kasus Merek Somethinc). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 481-489. <https://doi.org/10.57152/malcom.v4i2.1204>
- Putri, D. H., Santosa, I., Junaidy, D. W., & Adhitama, G. P. (2026). A Systematic Literature Review of Spatial Attributes and Well-Being in Women's Spa Experiences. *The International Journal of Architectonic, Spatial, and Environmental Design*, 20(1), 215.

- Putri, D. H., Santosa, I., Junaidy, D. W., & Adhitama, G. P. (2026). A Systematic Literature Review of Spatial Attributes and Well-Being in Women's Spa Experiences. *The International Journal of Architectonic, Spatial, and Environmental Design*, 20(1), 215.
- Rüdiger, M., Antons, D., Joshi, A. M., & Salge, T. O. (2022). Topic modeling revisited: New evidence on algorithm performance and quality metrics. *Plos one*, 17(4), e0266325. <https://doi.org/10.1371/journal.pone.0266325>
- Saparuddin, S., Kusumaningrum, A., Setiorini, A., & Takidah, E. (2025). The Rise Of The Beauty Industry In Indonesia: Assessing Its Contribution To National Economic Growth. *International Journal Of Economic Literature*, 2(10), 1013-1029.
- Syaputri, D. R., & Herawati, S. (2025). Influence Of Product Quality On Repurchase Intention Mediated By Customer Satisfaction On Sunscreen Products In Cimahi City. *BIMA Journal (Business, Management, & Accounting Journal)*, 6(2), 1377-1386. <https://doi.org/10.37638/bima.6.2.1377-1386>
- Vuković, D. B., Dekpo-Adza, S., & Matović, S. (2025). AI integration in financial services: a systematic review of trends and regulatory challenges. *Humanities and Social Sciences Communications*, 12(1), 562. <https://doi.org/10.1057/s41599-025-04850-8>
- Wang, S., Zhang, Y., Xiao, Y., & Liang, Z. (2025). Artificial intelligence policy frameworks in China, the European Union and the United States: An analysis based on structure topic model. *Technological Forecasting and Social Change*, 212, 123971. <https://doi.org/10.1016/j.techfore.2025.123971>
- Weston, S. J., Shryock, I., Light, R., & Fisher, P. A. (2023). Selecting the number and labels of topics in topic modeling: A tutorial. *Advances in Methods and Practices in Psychological Science*, 6(2), 25152459231160105. <https://doi.org/10.1177/25152459231160105>