

CHAPTER 1

INTRODUCTION

Online reviews from customers are a great way to learn about their preferences and how well products work [2]. However, it can be difficult to derive useful insights from the large amount of unstructured review data. In order to solve this, the data is methodically analyzed and distilled using advanced analytics techniques including topic modeling, social network analysis (SNA), and sentiment analysis. These techniques assist in determining important themes, investigating connections between product attributes, and eliciting customer feedback. This chapter outlines the essential components of the background of the study. This comprises the rationale, conceptual frameworks, statement of the problem, hypothesis and assumptions, scope and delimitation, and the importance of the study.

1.1 Rationale

Consumer preferences and product performance can be better understood by analyzing online customer reviews[2]. In the skincare sector, they have a direct effect on sales, brand loyalty, and customer happiness. While unfavorable reviews can drive customers to competitors, positive evaluations draw in new ones. In a market that is extremely competitive, satisfying consumer expectations through online reviews promotes loyalty, keeps a solid reputation, and increases sales[3].

The skincare market in Indonesia is expected to develop at a rate of 3.70 percent per year (CAGR 2024-2029) and reach USD 2.76 billion by 2024, according to Statista[4]. The sun protection segment is anticipated to contribute USD 180 million to this market, expanding at a quicker rate of 7.26 percent annually[5]. According to Reni Yanita, there were 1,039 cosmetic businesses by the end of 2023, up from 819 in 2021. This indicates that there are a lot of chances in the quickly growing cosmetics industry[6]. In this regard, maintaining competitiveness in an expanding market requires a thorough comprehension of customer feedback via sophisticated research.

Users can discover frank reviews of skincare products, including sunscreen, on the Indonesian beauty review website. Users can rate different items and discuss their experiences on this site, which generates a plethora of data for analysis. With hundreds of reviews at their disposal, it provides insightful information about how Indonesian consumers see various items and aids in determining the key features that are most important to them.

The term "key product features" describes particular qualities or traits of a product that customers most frequently discuss or find significant in their reviews[7]. These characteristics could include texture, absorption rate, aroma, skin compatibility, and cost when it comes to skincare products like sunscreens. Businesses must understand these essen-

tial characteristics, as they have a direct impact on consumer satisfaction and purchase decisions[8][9].

However, because it is impractical to read through all of the unstructured review content in depth, it is difficult to extract specific and actionable insights [10]. Thus, a multifaceted approach is required to distill this information into essential elements, including the primary subject of a comment, a viewpoint, or an evaluation of a product or service. Three main analytical methods are utilized to accomplish this: Topic Modeling, Social Network Analysis (SNA), and Sentiment Analysis.

Topic Modeling is an unsupervised machine learning technique that reveals hidden themes or topics in large text corpora [11]. Sentiment Analysis detects customer emotions and opinions expressed in text [12], offering valuable insight into consumer satisfaction and supporting informed decision-making [13]. Social Network Analysis (SNA) explores the semantic connections among features based on how users describe them in reviews.

In the study by Wang et al. (2020) [2], SNA was used to identify central product features in customer reviews. Their approach was based on entity analysis, which involved identifying co-occurring entities in sentences to construct a network of product aspects. Furthermore, Wang et al. emphasized the importance of incorporating sentiment analysis into such frameworks in order to capture not only what customers discuss but also how they feel about the identified features.

To adapt this framework to Indonesian-language reviews, where labeled entities are unavailable and entity recognition tools are limited, this study replaces entity analysis with Sentence Segment-LDA (SS-LDA) [14]. SS-LDA is an unsupervised topic modeling method derived from LDA [15] that is especially suited for short, unlabeled texts. It segments review content into aspect-focused units, allowing for the discovery of dominant product features without relying on manual annotations.

To evaluate customer perception of these features, this study employs Aspect-Based Sentiment Analysis (ABSA) [1, 16], using a hybrid method that combines lexicon-based sentiment scoring with machine learning classification. This approach allows sentiment to be directly associated with each extracted feature.

This study contributes by integrating three analytical methods: SS-LDA, ABSA, and SNA into a unified framework. SS-LDA enhances features detection by generating coherent segments in short-text reviews. ABSA enables fine-grained sentiment analysis by linking emotions to specific features. SNA adds a semantic layer by identifying how features are connected within customer discourse. Compared to previous studies that applied these techniques individually or in limited combinations, this integrated approach offers a more complete and actionable understanding of customer preferences. It supports both product development and strategic decision-making.

1.2 Conceptual Framework/Paradigm

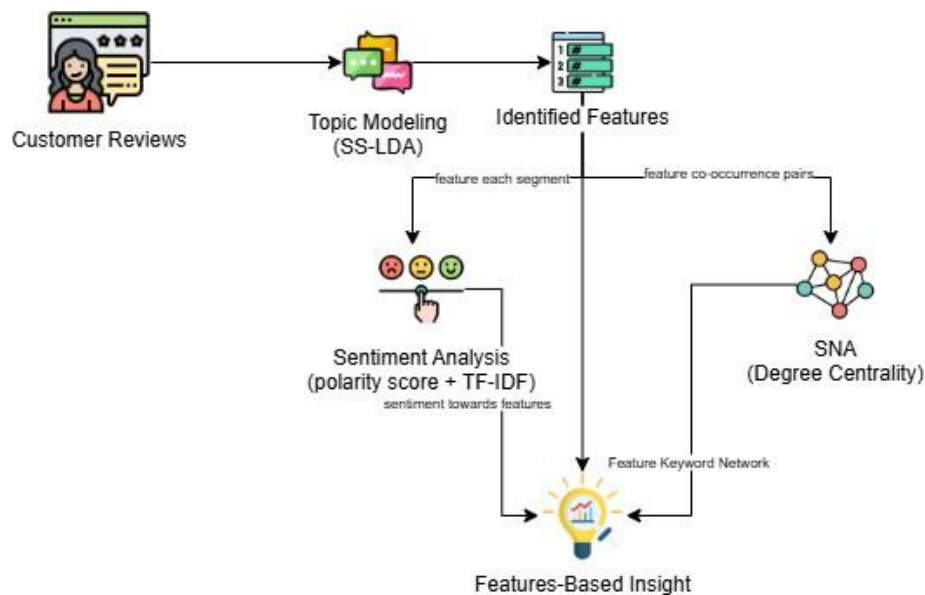


Figure 1.1: Methodological Relationship in Feature Analysis

Figure 1.1 illustrates the overall analytical flow used in this study. The process begins with customer reviews, which are the primary data source consisting of user-generated feedback on sunscreen products from the Female Daily platform, written in Bahasa Indonesia. These unstructured textual reviews are then processed using Sentence Segment-Latent Dirichlet Allocation (SS-LDA), a topic modeling technique specifically designed for short, unlabeled texts. For example, a review such as "Cepat menyerap, tapi bikin perih di mata" is segmented and analyzed to identify latent topics corresponding to features like texture and irritation.

From the SS-LDA results, identified features are derived from the top keywords that consistently represent important product attributes, such as "tekstur," "harga," or "SPF." These features become the foundation for two parallel analyses: sentiment analysis and network analysis. The sentiment analysis applies a hybrid method combining lexicon-based polarity scores with TF-IDF weighting to determine the emotional tone (positive, neutral, or negative) associated with each feature. For instance, terms like "harga murah" may indicate positive sentiment toward affordability, while "perih di mata" could reflect a negative experience related to irritation.

In parallel, a co-occurrence matrix is constructed using the probabilistic weights of top keywords within each topic, reflecting the joint probability of word pairs appearing in the same topic. These pairs, such as "harga" and "murah," represent semantically associated feature relationships. A semantic network is then built from this matrix and analyzed using Social Network Analysis (SNA) to calculate degree centrality, which quantifies the

centrality of each feature based on its weighted connections to others. Features with high degree centrality are considered more influential due to their strong and frequent associations with other product features.

The final stage synthesizes the insights from topic modeling, Aspect-Based Sentiment Analysis (ABSA), and SNA into a features-based insight. This output provides a comprehensive understanding of customer perception by integrating what features are discussed, how customers feel about them, and how those features are interrelated in customer discourse. Although each analytical component is executed independently, they are logically connected, enabling a multidimensional interpretation of user-generated content that can inform product development and strategic decision-making.

1.3 Statement of the Problem

The growing volume of customer reviews in the skincare industry presents valuable opportunities for extracting product insights. However, the unstructured and diverse nature of these reviews makes it difficult to extract specific and actionable information, especially when written in Indonesian without standardized aspect annotations. Traditional metrics like star ratings and sales volumes provide limited insights into specific consumer preferences or dissatisfaction points [2].

Although previous studies, such as Wang et al. [2], have employed Social Network Analysis (SNA) based on entity analysis to identify product features and their interrelations, such methods rely on the availability of annotated entities and robust language models, which are limited in the Indonesian context. Additionally, these methods often lack integration with sentiment analysis, thereby missing the emotional dimension of customer perception.

Therefore, there is a need for a comprehensive analytical framework that can automatically identify key product features, assess sentiment associated with those features, and explore their semantic relationships within customer reviews. This study addresses this gap by integrating Sentence Segment-LDA (SS-LDA), Aspect-Based Sentiment Analysis (ABSA), and Social Network Analysis (SNA) to analyze unstructured sunscreen reviews.

This integrated framework provides a multidimensional understanding of consumer perception by bridging feature identification, sentiment evaluation, and semantic relationships, ultimately enabling businesses to enhance product development and align with consumer expectations.

1.4 Objective and Hypotheses

1.4.1 Objective

The main objective of this study is to design and implement an analytical framework that integrates Sentence Segment-LDA (SS-LDA), Aspect-Based Sentiment Analysis (ABSA), and Social Network Analysis (SNA) to explore key product features mentioned in customer reviews of sunscreen products on the Female Daily platform. Specifically, this study aims to extract frequently mentioned product features using SS-LDA, assess customer sentiment toward each feature through ABSA, analyze the relationships among features based on their co-occurrence patterns among frequently mentioned terms using Social Network Analysis (SNA) with degree centrality.

1.4.2 Hypotheses

This research hypothesizes that the integration of topic modeling, sentiment analysis, and social network analysis (SNA) will provide comprehensive insight into key product features. Specifically, This study hypothesizes that Sentence Segment-LDA (SS-LDA) is effective in extracting meaningful and interpretable topics that represent key product features frequently discussed by customers, Aspect-Based Sentiment Analysis (ABSA) can accurately capture the sentiment polarity (positive, neutral, or negative) toward specific product features mentioned in customer reviews, degree centrality in Social Network Analysis (SNA) highlights the most influential product features by revealing co-occurrence patterns and connections among frequently mentioned words, and The combined findings from topic modeling, sentiment analysis, and network analysis can provide useful insights for skincare brands to improve product development and marketing strategies.

1.5 Assumption

This study assumes that customer reviews on the Female Daily platform authentically represent the true opinions and experiences of customers concerning sunscreen goods. Furthermore, it is anticipated that utilizing data in the Indonesian language will not hinder the capacity to derive significant conclusions. These assumptions establish the basis for confirming that the data utilized in this study is appropriate for subsequent analysis.

1.6 Scope and Delimitation

This study examines customer reviews of the most popular sunscreen products on the Female Daily website. The principal objective is to identify key product features, evaluate customer sentiment, and analyze the interconnections among product features.

- Topic Modeling (SS-LDA) is employed to generate topics representing key product features. These topics, such as Texture, SPF Protection, and Affordability, are extracted from customer reviews to understand the most discussed product attributes.
- Sentiment Analysis (ABSA) is conducted to determine the sentiment polarity (positive, neutral, or negative) of each topic, providing insights into how customers perceive different product attributes.
- Social Network Analysis (SNA) is limited to degree centrality analysis on top keywords extracted from the topic modeling results.
- The dataset comprises only text-based reviews, restricted to those from the year 2024 and focuses on the most popular sunscreen products reviewed on Female Daily in 2024.
- Expert validation is performed using a questionnaire to ascertain if the findings yield significant insights for product enhancement.
- This study does not include the development of a system or interactive dashboard. The final output is limited to the analytical results of each method through visualizations and descriptive analysis.

1.7 Significance of the Study

This research offers significant insights for both scholars and industry professionals. This research employs Topic Modeling, Sentiment Analysis, and Social Network Analysis to examine consumer reviews of popular sunscreen products, thereby identifying essential product attributes and evaluating user sentiment. These findings can aid skincare brands in comprehending consumer preferences, hence informing product development and marketing strategies. This study enhances the academic field by illustrating the efficacy of integrated analytical methodologies in deriving significant insights from unstructured data, specifically within the Indonesian skincare market.