Jurnal Infact Sains dan Komputer
Vol.10, No.01, May 2026
p-ISSN : 2527-8363, e-ISSN: 2829-0259
DOI:

Implementation of Latent Dirichlet Allocation Topic Modeling and VADER on Aspect-Based Sentiment Analysis

Kevin Miracle Satoko[1], Sunneng Sandino Berutu[2]*, Jatmika[3], Retno Palupi[4]
[1,2]Program Studi Informatika, Fakultas Sains dan Komputer, Universitas Kristen Immanuel Yogyakarta
Jln. Ukrim No. KM 11, Kadirojo I, Purwomartani, Kalasan, Sleman, Daerah Istimewa Yogyakarta 55571, Indonesia
[3]Program StudiTeknikInformatika, Fakultas Teknik, Universitas Kristen Teknologi Solo
Jl.R.W. Monginsidi No. 36-38 Surakarta, Indonesia

E-mail: kevin.m19@student.ukrimuniversity.ac.id[1],
sandinoberutu@gmail.com[2]*, jatmika@ukrimuniversity.ac.id.com[3], palupiretno748@gmail.com[4]
Submitted: 03/04/2025, Revision: 04/22/2026, Accepted : 05/07/2026

Abstract

Aspect-Based Sentiment Analysis on a Product or Service is Crucial for Enhancing Customer Satisfaction. This Study Applies Latent Dirichlet Allocation (LDA) Topic Modeling to Identify Aspects. Then, the Valence Aware Dictionary and Sentiment Reasoner (VADER) Lexicon Method is Adopted to Determine Sentiment on These Aspects. The Data Source Comes from Customer Reviews of a Gelato Ice Cream Shop at Taman Siswa. Data was collected from Google Maps Using the Web Scraping Method via the Instant Data Scrapper Application. The Experimental Results Show that the LDA Method Identified 3 Aspects: Flavor, Place, and Service. Meanwhile, Sentiment Measurement Using VADER on the Flavor Aspect Revealed a Positive Sentiment of 213%, Negative Sentiment of 60%, and Neutral Sentiment of 218%. The Next Aspect, Place, Had a Positive Sentiment of 32%, Negative Sentiment of 4%, and Neutral Sentiment of 4%, while the Service Aspect Composed of 32% Positive Sentiment, 3% Negative Sentiment, and 3% Neutral Sentiment. Overall, the Positive Sentiment for the Flavor Aspect (Taste) Outweighed the Negative and Neutral Sentiments for the Place (Location) and Service (Service) Aspects.

Keywords: LDA,;VADER;aspek; sentiment; gelato;

***Corresponding author: Kevin miracle satoko**
E-mail: kevin.m19@student.ukrimuniversity.ac.id

1. Introduction

In the rapidly growing digital era, user reviews can be managed to understand customer perceptions and satisfaction with products or services. Tempo Gelato, a popular destination in Yogyakarta, serves as a real-world case study where customer reviews significantly influence the marketing image of the Tempo Gelato outlet. Understanding and manually analyzing customer reviews can be a highly complex task, given the large volume of data and the variety of opinions expressed by consumers. Therefore, by performing aspect-based sentiment analysis, the LDA and VADER methods are applied to analyze these reviews, both in terms of aspects and the sentiment contained within them. In general, the LDA method is implemented in topic modeling. In the context of review analysis, LDA can be used to group words in reviews into specific topics or aspects. By combining LDA with sentiment analysis, deeper insights into customer reviews can be gained, not only by determining whether a review is positive or negative but also by understanding which aspects customers like or dislike. LDA is a three-tier hierarchical Bayesian model, where each item from a collection is modeled as a finite mixture of a set of underlying topics. Each topic, in turn, is modeled as an infinite mixture of a set of underlying topic probabilities. In text modeling, topic probabilities provide an explicit representation of a document [1]. VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool specifically designed to handle the nuances of social media text. Developed by C.J. Hutto and Eric Gilbert in 2014, VADER is unique due to its ability to accurately analyze sentiment in short and informal texts, such as tweets, comments, and reviews [2]. Research also shows that VADER outperforms other similar lexicon-based methods, such as TextBlob [3]. This study aims to identify aspects through topic modeling using the LDA method and measure the positive, neutral, and negative sentiments on these aspects with VADER.

1. Research Methods

The dataset was taken from Google Maps reviews. Following the data extraction, it underwent a preprocessing stage. After that, it was processed through topic modeling with LDA and evaluated using the Coherence Score. The results of the modeling serve as a reference for determining the aspects. Finally, the sentiment of each aspect was measured using the VADER method. The research stages are presented in Figure 1.1.

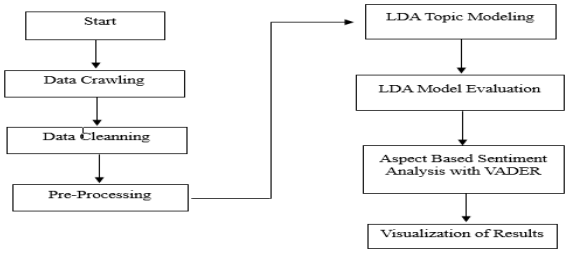

Figure 1.1 Research Stages

1.1 Data Collection

Data was collected using scraping techniques with the help of the additional application Instant Data Scrapper. After the data was obtained, it was stored in a CSV file.

2.2 WordCloud

WordCloud was used to visualize words based on their frequency of occurrence, with word visualization featuring different font sizes and colors [4].

2.3 Pre-Processing Data

Preprocessing was the most important stage before topic modeling, as it involved text cleaning and the removal of irrelevant text characters to ensure cleanliness and ease in processing data at the next stage. The preprocessing steps included: cleaning text, case folding, normalization, tokenization, stopword removal, and stemming. In the cleaning text stage, irrelevant characters such as usernames, blank spaces, special characters, and URLs were removed [5]. The case folding was done by converting all text to lowercase and removing irrelevant punctuation [6]. The next stage was tokenization, which involved grouping sentences into sequences of words [7], followed by stopword removal to speed up the analysis [8]. Finally, lemmatization was used to reduce inflected words to their base forms [9].

2. Results and Discussion

3.1 Crawling data

The number of reviews obtained was 360 data points with the help of the Instant Data Scrapper application. The scraping results were stored in .csv format. This dataset consists of two columns: 'Name' and 'Review'. A sample of the collected data can be seen in the image below.

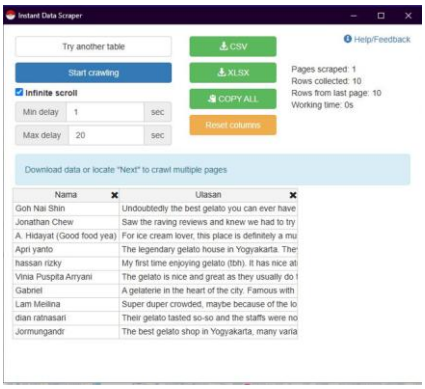

Figure 3.1 Sample of Data Scraper Results

3.2 Pre-Processing Result

The results of the preprocessing stage of the dataset are presented in the following table, showing a snippet of some words.

3.2.1 Cleaning Text Result

Table 3.2.1 Cleaned text

REVIEW DATA	CLEANED TEXT
Undoubtedly the best gelato you can ever have ...	undoubtedly the best gelato you can ever have ...
Saw the raving reviews and knew we had to try ...	saw the raving reviews and knew we had to try ...
For ice cream lovers this place is definitely ...	for ice cream lovers this place is definitely ...
The legendary gelato house in Yogyakarta They ...	the legendary gelato house in Yogyakarta they ...
My first time enjoying gelato tbh It has nice ...	my first time enjoying gelato tbh it has nice ...

3.2.2 Tokenization

Table 3.2.2 *Tokenization*

REVIEW DATA	TOKENIZATION
Undoubtedly the best gelato you can ever have ...	[undoubtedly, the, best, gelato, you, can, ever, have, ...
Saw the raving reviews and knew we had to try ...	[saw, the, raving, reviews, and, knew, we, had, to, try, ...
For ice cream lovers this place is definitely ...	[for, ice, cream, lover, this, place, is, definitely, ...
The legendary gelato house in Yogyakarta They ...	[the legendary, gelato, house, in, Yogyakarta, they, ...
My first time enjoying gelato tbh It has nice ...	[my, first, time, enjoying, gelato, tbh, it, has, nice, ...

3.2.3 Removing Stopwords

Table 3.2.3 Removing Stopwords

REVIEW DATA	REMOVING STOPWORDS
Undoubtedly the best gelato you can ever have ...	undoubtedly, the, best, gelato, you, can, ever, have, ...
Saw the raving reviews and knew we had to try ...	saw, the, raving, reviews, and, knew, we, had, to, try, ...
For ice cream lovers this place is definitely ...	for, ice, cream, lover, this, place, is, definitely, ...
The legendary gelato house in Yogyakarta They ...	the, legendary, gelato, house, in, yogyakarta, they, ...
My first time enjoying gelato tbh It has nice ...	my, first, time, enjoying, gelato, tbh, it, has, nice, ...

3.2.4 Lemmatization

Table 3.2.4 Lemmatization

REVIEW DATA	LEMMATIZATION
Undoubtedly the best gelato you can ever have ...	undoubtedly, the, best, gelato, you, can, ever, have, ...
Saw the raving reviews and knew we had to try ...	saw, the, raving, reviews, and, knew, we, had, to, try, ...
For ice cream lovers this place is definitely ...	for, ice, cream, lover, this, place, is, definitely, ...
The legendary gelato house in Yogyakarta They ...	the, legendary, gelato, house, in, Yogyakarta, they, ...
My first time enjoying gelato tbh It has nice ...	my, first, time, enjoying, gelato, tbh, it, has, nice, ...

3.2.5 Stemming

Table 3.2.5 Stemming

REVIEW DATA	STEMMING
Undoubtedly the best gelato you can ever have ...	undoubtedly, the, best, gelato, you, can, ever, have, ...
Saw the raving reviews and knew we had to try ...	saw, the, raving, reviews, and, knew, we, had, to, try, ...
For ice cream lover this place is definitely ...	for, ice, cream, lover, this, place, is, definitely, ...
The legendary gelato house in Yogyakarta They ...	the, legendary, gelato, house, in, Yogyakarta, they, ...
My first time enjoying gelato tbh It has nice ...	my, first, time, enjoying, gelato, tbh, it, has, nice, ...

3.3 WordCloud

3.3.1 WordCloud Dataset Before Preprocessing

The word cloud of the dataset before preprocessing is visualized in Figure 3. The words 'place' and 'gelato' have the largest font size compared to the other words.

Figure 3.3.1 Wordcloud Before Preprocessing

3.3.2 WordCloud Dataset After Preprocessing

The word cloud of the dataset after preprocessing is visualized in Figure 4. The visualization results show that the words 'place' and 'gelato' have a large and prominent font size compared to the other words.

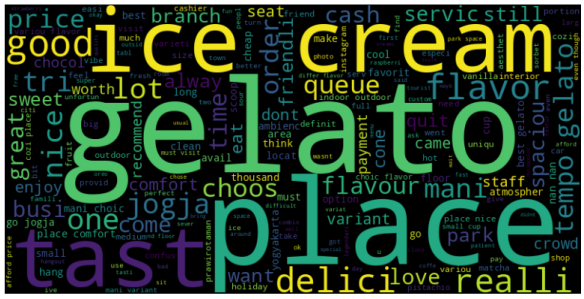

Figure 3.3.2 Wordcloud After Preprocessing

3.4 LDA Topic Modeling

3.4.1 Latent Dirichlet Allocation

In this process, an LDA model was created using the input dictionary and corpus [10]. The results of the topic modeling with LDA identified 5 relevant topics, each with 10 keywords and their respective scores. The results of the LDA topics can be seen in Table 3.4.1 below.

Table 3.4.1 Word Probability

Topics	Probability
0	('0.027*"place" + 0.022*"ice" + 0.020*"cream" + 0.019*"nice" + 0.016*"gelato" + 0.015*"one" + 0.011*"jogja" + 0.010*"good" + 0.009*"order" + 0.009*"tast"')
1	('0.027*"place" + 0.016*"ice" + 0.016*"cream" + 0.015*"flavor" + 0.015*"jogja" + 0.013*"nan" + 0.013*"lot" + 0.012*"gelato" + 0.012*"park" + 0.011*"nice"')
2	('0.036*"gelato" + 0.031*"place" + 0.011*"best" + 0.011*"good" + 0.009*"mani" + 0.008*"time" + 0.008*"great" + 0.008*"one" + 0.008*"realli" + 0.008*"flavour"')
3	('0.034*"gelato" + 0.027*"queue" + 0.021*"flavor" + 0.018*"long" + 0.014*"delici" + 0.012*"busi" + 0.010*"good" + 0.009*"come" + 0.009*"still" + 0.008*"realli"')
4	('0.046*"gelato" + 0.040*"place" + 0.024*"tast" + 0.022*"ice" + 0.021*"cream" + 0.019*"flavor" + 0.016*"delici" + 0.015*"realli" + 0.014*"mani" + 0.013*"good"')

3.4.2 Aspect Extraction

The aspect extraction stage was done manually to determine the aspects for each review based on the dominant topics. Data labeling was then performed by assigning sentiment labels to each aspect. At this stage, the topics were linked to the aspects of Place, Flavor, and Service. The following table shows the aspect extraction based on the dominant topics generated by LDA.

Table 3.4.2 Aspect Extraction

Topics	***Relevant words/keywords***	***Aspect***
0	*"place", "ice", "cream", "nice", "gelato", "good", "order", "taste"*	*PLACE*
1	*"place", "ice", "cream", "flavor", "gelato", "park"*	*FLAVOR*
2	*"gelato", "place", "best", "good", "great", "time", "flavour"*	*FLAVOR*
3	*"gelato", "queue", "long", "delicious", "busy", "come"*	*SERVICE*
4	*"gelato", "place", "taste", "ice", "cream", "delicious"*	*FLAVOR*

3.4.3 Aspect Labeling

The labeling aspects stage based on the extraction results generated by the LDA model resulted in sentiment aspects for each review document.

Table 3.4.3 Aspect Labeling

Text	***Dominant topic***	***Aspect***
undoubtedli best gelato ever indonesia probabl...	4	*Flavor*
saw rave review knew tri inde didnt disappoint...	4	*Flavor*
ice cream lover place definit must visit one v...	0	*Place*
legendari gelato hous yogyakarta variou flavor...	4	*Flavor*
first time enjoy gelato tbh nice atmoshpher en...	4	*Flavor*

3.5 LDA Model Evaluation

The results of the testing were used for the CV coherence method. Figure 4.1 shows the evaluation of the LDA model using the Coherence Score to

determine the optimal number of topics. The following is the analysis:

- X-axis (Number of Topics): The number of topics tested (from 2 to 10).
- Y-axis (Coherence Score): The coherence score for each number of topics, which measures the quality of topics based on the uniformity of words within each topic. A higher value indicates that the topic is more meaningful and coherent.

Trend:
The highest coherence was achieved at 3 topics (around 0.38). After that, the coherence tended to decrease as the number of topics increased, although there was a slight increase at 8 topics.

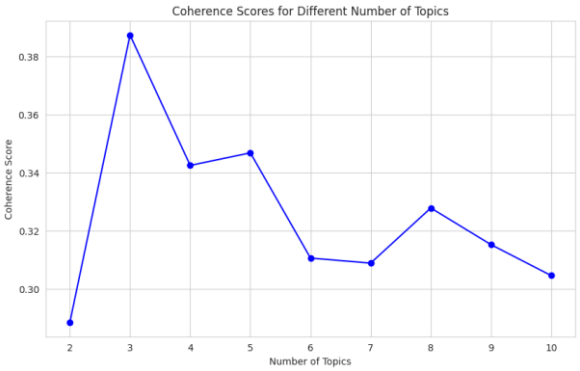

Figure 4.1 Model Evaluation (coherence score)

3.6 Aspect-Based Sentiment Analysis *(VADER)*

The following determines aspect-based sentiment analysis using the Valence Aware Dictionary and Sentiment Reasoner (VADER). The results can be seen in Table 3.6 below.

Table 3.6 Sentiment Aspect

Text	*Dominan topic*	*Aspects*	*Sentiment*
undoubtedli best gelato ever indonesia probabl...	*4*	*FLAVOR*	*Positif*
saw rave review knew tri inde didnt disappoint...	*4*	*FLAVOR*	*Positif*
ice cream lover place definit must visit one v...	*0*	*PLACE*	*Positif*
legendari gelato hous yogyakarta variou flavor...	*4*	*FLAVOR*	*Positif*
first time enjoy gelato tbh nice atmoshpher en...	*0*	*PLACE*	*Positif*

3.7 Results Visualization

3.7.1 *pyLDAvis*

Figure 3.7.1 displays the results of the Latent Dirichlet Allocation (LDA) model visualization, which is visualized using an interactive tool commonly used to interpret LDA results.

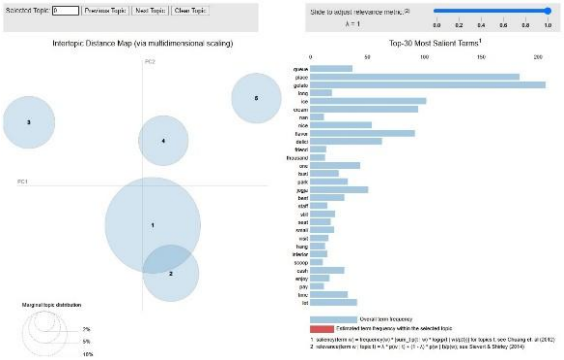

Figure 3.7.1 The Result of pyLDAvis

3.7.2 Diagram Batang

Figure 3.7.2 shows the sentiment distribution (positive, negative, neutral) based on the aspects evaluated, namely FLAVOR, PLACE, and SERVICE.

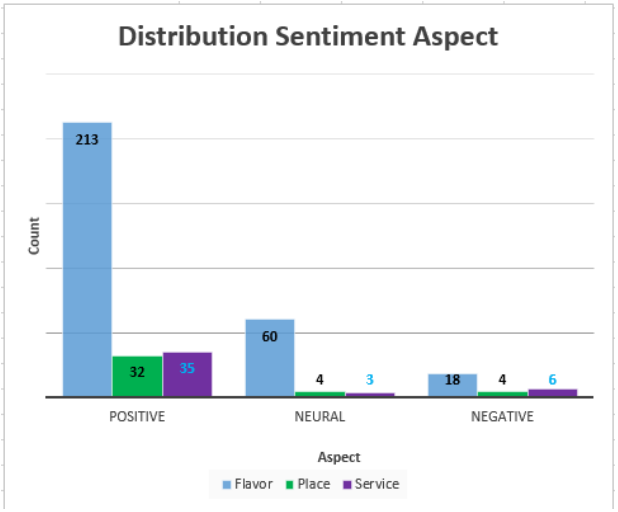

Figure 3.7.2 Aspect Sentiment Distribution

3.7.3 WordCloud

This WordCloud represents the keywords of the topic with the highest coherence, as indicated by the title 'WordCloud of keywords from the topic with the highest coherence.' In the WordCloud, the size of each word is generally proportional to its frequency or importance in the data.

Figure 3.7.3 Highest Coherence Wordcloud

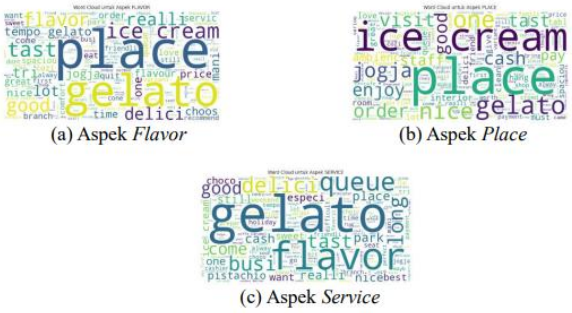

Figure 3.11 WordCloud for Each Aspect

3. Conclusions

Based on the results of the research conducted, several conclusions can be explained as follows:

- Through the application of the Topic Modeling method, using the Latent Dirichlet Allocation (LDA) technique, 5 topics were generated, each with 10 keywords. The 5 topics with the highest coherence scores were identified, aspect extraction was performed, and it was successful.
- Sentiment analysis was processed using the Valence Aware Dictionary and Sentiment Reasoner (VADER) method. In its implementation, it successfully provided aspect-based sentiment analysis for the 3 main aspects. The Flavor aspect resulted in 213 positive review data, 60 neutral review data, and 18 negative review data. The Place aspect resulted in 32 positive review data, 4 neutral review data, and 4 negative reviews, while the Service aspect resulted in 35 positive review data, 3 neutral review data, and 6 negative review data. Based on the sentiment analysis data, the aspect with the most positive reviews was Flavor, so it can be concluded that customers of Tempo Gelato Taman Siswa are very satisfied with the flavor of Tempo Gelato Taman Siswa's ice cream.

References:

[1] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." Journal of machine Learning research 3.Jan (2003): 993-1022.

[2] Youvan, Douglas C. "Understanding sentiment analysis with VADER: a comprehensive overview and application." AI and Data Science Journal (2024).

[3] Anwar, Muchamad Taufiq. "Analisis Sentimen Masyarakat Indonesia Terhadap Produk Kendaraan Listrik Menggunakan VADER." JATISI (Jurnal Teknik Informatika dan Sistem Informasi) 10.1 (2023): 783-792.

[4] Permadi, V. A. (2020). Analisis sentimen menggunakan algoritma Naive Bayes terhadap review restoran diSingapura. Jurnal Buana Informatika, 11(2), 141. Retrieved from https://ojs.uajy.ac.id/index.php/jbi/article/view/3769.

[5] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," JST (Jurnal Sains dan Teknologi), vol. 12, no. 1, pp. 129–137, Apr. 2023, doi: 10.23887/jstundiksha.v12i1.52358.

[6] Y. Alotaibi et al., "Suggestion mining from opinionated text of big social media data," Computers, Materials and Continua, vol. 68, no. 3, 2021, doi: 10.32604/cmc.2021.016727.

[7] Syaifulloh Amien Pandega Perdana, Teguh Bharata Aji, and Ridi Ferdiana, "Aspect Category Classification dengan Pendekatan Machine Learning Menggunakan Dataset Bahasa Indonesia," Jurnal Nasional Teknik Elektro dan Teknologi Informasi, vol. 10, no. 3, 2021, doi: 10.22146/jnteti.v10i3.1819.

[8] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," Expert Syst Appl, vol. 212, 2023, doi: 10.1016/j.eswa.2022.118715.

[9] N. Z. Dina and N. Juniarta, "Aspect based Sentiment Analysis of Employee's Review Experience," Journal of Information Systems Engineering and Business Intelligence, vol. 6, no. 1, 2020, doi: 10.20473/jisebi.6.1.79-88.

[10] Sagala, Hetthroh, and Hapnes Toba. "Penentuan Aspek yang Berpengaruh Terhadap Produk Smartphone Berdasarkan Ulasan Berbasis Tekstual." *Jurnal Teknik Informatika dan Sistem Informasi* 7.1 (2021).