CENTIVE
Conference on Electrical Engineering, Informatics,
Industrial Technology, and Creative Media

ISBN: 978-602-53004-0-0

A Lexicon-Based VADER Approach for Aspect-Based Sentiment Analysis in the Indonesian Language

Muhammad Arsy Al Banna [1], Siti Khomsah*[1]
[1]Data Scicence Study Program, Telkom University, Purwokerto Campus
[1]Jl.DI Pandjaitan 128, Banyumas 53147, Indonesia
*sitijk@telkomuniversity.ac.id

Abstract — Aspect-Based Sentiment Analysis (ABSA) provides detailed insights into customer opinions by identifying specific aspects—such as product, service, and management—in textual reviews and analyzing the sentiment toward each aspect. Unlike general sentiment analysis, ABSA reveals which dimensions of customer experience require improvement. However, applying ABSA in low-resource languages like Indonesian is challenging due to limited annotated dataset, sentiment lexicon, and pre-trained model, which often reduce the accuracy of machine learning or deep learning approaches. This study employs the Valence Aware Dictionary for Sentiment Reasoning (VADER), a lexicon-based algorithm effective in analyzing short, informal, and mixed-language texts, such as online reviews. VADER enables reliable sentiment scoring without large labeled datasets, making it suitable for Indonesian-language analysis. A total of 8,438 Google Maps reviews from 2016 to 2025 were analyzed to observe sentiment trends over time. Keywords were developed for three main aspects:product (1,112 words), service (468 words), and management (666 words). Results show that most reviews express positive sentiment (85.4%), followed by neutral (9.9%) and negative (4.6%). The product aspect was most discussed (7,839 reviews), followed by management (4,608) and service (4,589). In conclusion, VADER-based ABSA can effectively analyze customer sentiment in low-resource languages, providing actionable insights to guide restaurant service improvements. The lack of VADER are an obstacle in handling nuance in the Indonesian language and many keywords cannot be extracted by VADER. Futher, method development is needed for more precise aspect extraction.

Keywords – ABSA, VADER, Sentiment Analysis, Indonesian language.

I. INTRODUCTION

The restaurant industry is a commercial sector with significant economic potential as it fulfills one of the primary human needs - food. The advancement of digital technology, particularly through platforms such as Google Maps, has transformed the competitive landscape by allowing customers to share reviews that directly influence a restaurant's reputation and brand image. These reviews can be analyzed using sentiment analysis, a method for identifying and categorizing customer opinions to obtain structured insights into their perceptions of product, service, and management aspects. According to SNI 01-6175-2014 and the Regulation of the Minister of Tourism and Creative Economy No. 11 of 2014 [1], restaurants are required to meet standardized criteria in these three aspects. Restaurant owners can utilize big data analytics, such as sentiment analysis derived from online review datasets, to understand sentiment polarization—the variation between positive and negative opinions expressed by customers. This context is particularly relevant for restaurants in Banyumas Regency, which face challenges arising from differing preferences between local residents and tourists, as well as increasing competition due to the emergence of new restaurants. These conditions require restaurant managers to gain deeper insights into customer perceptions to improve the quality of their products and services. Therefore, this study aims to analyze customer sentiment polarization toward restaurants in Banyumas Disctrict based on standardized aspects of product, service, and management, in order to evaluate their implementation and support quality enhancement in the local restaurant industry. Usually, customer opiniom triggered by specificts aspect such as services, price, quality of products etc [2], [3]. The difficulty to dig aspect in sentiment analysis is classify senetences into aspects [4]. Several studies have explored aspect-based sentiment analysis (ABSA) to understand customer opinions of restaurans.

Parasati, Abdurahman Bachtiar, and Setiawan [5] developed an ABSA model for restaurant reviews using Naive Bayes with TF–IDF feature extraction on TripAdvisor and Google Reviews, achieving accuracies of 88%, 84%, and 76% for food, atmosphere, and service aspects, respectively. Similarly, Nofandi, Setyawan, and Brata[6] employed Support Vector Machine (SVM) combined with root cause analysis to identify major issues in restaurant reviews, revealing key problems related to food quality, pricing, and service efficiency. In both studies, aspect labeling and sentiment annotation were performed manually, which limited scalability and automation.

Subsequent research extended ABSA to other domains such as finance services [7], public transportation [8]. The data sources crawled from social media platforms such as Twiter or public platform such as Google Maps [9], [10]. Yutika, Adiwijaya, and Faraby [11]translated English aspects into Indonesian using Complement Na¨ıve Bayes, while Fadillah, Hamami, and Andreswari [7] , and Radiena and Nugroho [8] applied SVM and Random Forest for tourism and hospitality datasets, achieving accuracies between 74% and 90%.

Although those approaches yielded strong performance, most relied on manual annotation for aspect and sentiment labeling and required substantial preprocessing and training data. This dependency makes them less adaptable to dynamic review data and limits real-time sentiment evaluation. To overcome these limitations, the present study introduces a semi-automated ABSA framework usingnthe Valence Aware Dictionary and sEntiment Reasoner (VADER). VADER is a lexicon and rulenbased sentiment analysis algorithm optimized for social media and online reviews, capable of detecting sentiment intensity and contextual polarity without manual annotation or large labeled datasets [12], [13], [14]. By integrating aspect lexicons with VADER's sentiment scoring, this research aims to automate aspect extraction and sentiment classification on restaurant reviews in Banyumas Regency. This approach enhances efficiency, scalability, and inter- pretability, addressing the major limitations of previous ABSA studies.

II. RESEARCH METHOD

The research was conducted on restaurants in Banyumas Regency, with the subjects being the restaurants and the objects being customer sentiments and their influencing aspects. The workflow includes data crawling, data cleaning, and exploratory analysis, followed by sentiment labeling and aspect extraction using the VADER algorithm.

A. Data Crawling

First, Data were collected from Google Maps reviews, selecting only restaurants with a minimum of 100 reviews that were still running in 2025. From 154 restaurants, 71 met these criteria, and their reviews spanning 2016 to 2025. Data were scraped manually and verified to include personal usernames, relevant aspects, and exclude promotional or spam content.

B. Data Preprocessing and Exploratory Data Analysis

The next step is preprocessing to prepare the data for analysis. This preprocessing stage is conducted without tokenizing or stemming to maintain consistency with the VADER method's characteristics. The purpose of this process is to clean, simplify, and standardize the text format to achieve more accurate analysis results. At this stage, cleaning and letter normalization are performed on customer reviews. Removes irrelevant characters such as @, #, URLs, numbers, non-ASCII characters, and punctuation symbols, eliminates duplicate characters, and replaces line breaks with spaces using regular expressions. Subsequently, the casefolding task converts all text to lowercase to ensure consistency during token matching and other text processing. After cleaning, the average text length decreased by 3.9%, indicating that most of the data were already relatively clean - requiring only minor removal of irrelevant characters.

Text Normalization- this stage involves replacing non-standard words with their formal forms using a slang dictionary containing 937 entries. Subsequently, text normalization is performed by converting slang words into their standard equivalents. The normalization process using a slangword dictionary, which was built by own. The process checks each word in the reviews and replaces slang words if a corresponding entry exists in the dictionary. Words without matches are left unchanged. Next, stopwords—common words that carry little analytical meaning, such as conjunctions and prepositions are removed. We apply the NLTK library to do this task. We filter each word in the reviews, keeping only words that are not included in the stopword list.

After preprocessing, the next stage is conducting EDA to gain deeper insights into the content of the review data and to identify words that are irrelevant or disruptive in the aspect extraction process. This stage focuses on two main areas: noise detection and word frequency per aspect.

1) *Noise Detection*

Handling text noise in sentiment analysis requires caution. Sometimes a word considered noise can contain sentiment. Standard methods are not necessarily appropriate for the dataset being processed, so it is necessary to choose the right method [15], [16], [17][18].

This research apply word frequency analysis to reveal highfrequency words with minimal relevance to aspect determination. Words such as "nya," "iya," "suka," "anak," and "sih" fall into this category. This process involves concatenating all entries into a single large string and counting the frequency of each word using the counter function. Noise words are then identified based on the criteria of appearing more than 10 times but having a length of four characters or less. These words are neutral, expressive, or do not convey any specific topical meaning. To simplify understanding of the distribution of noise words, a graph visualization of the top 30 noise words based on their frequency is shown by Figure 1.

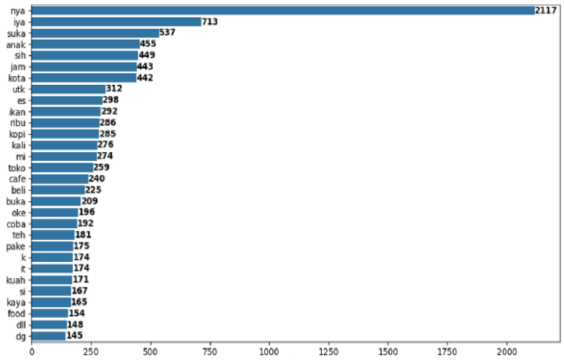

Fig. 1 The 30 words appear most frequently

Overall, a total of 258 words were identified as noise words. Identifying and filtering these words is crucial to prevent them from interfering with the aspect classification process in the subsequent analysis stages. As in Figure 1, there are 30 words that appear most frequently. These words appear to be typos or abbreviations.

2) Sentiment Classification

In this study, sentiment labeling was extracted using the VADER- a lexicon-based sentiment analysis model capable of determining data diversity. VADER's lexicon data consists of 7,500 words and 3,570 emojis, this condition is because VADER uses a human-centric approach which is a combination of qualitative analysis and empirical validation according to human judgment through emotional intensity [19]. The VADER method was developed by George Berry, Ewan Klein and Pier Paolo under the MIT license [12]. The VADER approach uses each word in the lexicon through an assessment process whether the word is positive, negative or neutral using a composite score.

This research using 3 class sentiments: positive, neutral, and negative to classify review. VADER employs formula compound score to determine class sentiment, as calculated in formula (1)[20].

$$\text{Compound score} = \frac{x}{\sqrt{x^2 + a}} \tag{1}$$

where x is Sum of the sentiment scores of the words in the sentence α is normalization parameter (default value = 15).

The compound score for each review represents the overall polarity of the text. Sentiment labels were then assigned according to predefined thresholds. This approach enables quantitative analysis of customer perceptions of restaurants based on the opinions expressed in their reviews. The pseudo code for sentiment annotation is below.

```
def analyze_sentiment(review):
    if not isinstance(review, str) or review.strip() == "":
        return 0.0, "Neutral"
    sentiment_scores = sent_analyzer.polarity_scores(review)
    compound_score = sentiment_scores["compound"]
    if compound_score >= 0.05:
        sentiment_label = "Positive"
    elif compound_score <= -0.05:
        sentiment_label = "Negative"
    else:
        sentiment_label = "Neutral"
    return compound_score, sentiment_label

# function for calculate compound score
df[["Compound_Score", "Sentiment"]] =
df["Description_Translate_en"].apply(analyze_sentiment).apply(pd.Series)
df.to_excel("Aspect_and_VADER_Sentiment_Results.xlsx", index=False)
```

C. *Aspect Extraction*

Aspect was extracted based on the components and subcomponents listed that established by the Ministry of Tourism and Creative Economy of the Republic of Indonesia[1]. These standards cover essential elements to ensure quality in restaurant operations. Keyword selection is enriched through common terms in customer reviews and data exploration results, such as the frequency of occurrence of relevant words in each aspect. This selection aims to capture terms directly used by consumers in describing their restaurant experiences. This keyword's list is dynamic and can be adjusted if new term discovered from customer review data. Examples of the results of component and sub component listed in Tables 1, 2, and 3 below.

Table 1 Components and Subcomponents of Aspect of Product

Component Aspect Product	***Sub Component of Aspect***
Dining area and circulation	Dining Room, Circulation
Provision of food and beverages.	Menu and recipies
Supporting facilities.	Waiting room, toilet, first aid kit, fire extinguisher, menu list and parking.

Table 2 Components and Subcomponents of Aspect of Services

Component Aspect of Services	***Sub Component of Aspect***
Standard operating procedures	Waiter, Payment, Cleanliness.
Other facilities	Places of worship, Wifi.

Table 3 Components and Subcomponents of Aspect of Management

Component Aspect Management	***Sub Component of Aspect***
Organization	Standard operating procedures
Management	Product development, hygiene, sanitation, safety, and health products
Human resources	Neatness of employee attire
Facilities and infrastructure	Employee facilities, environment, emergency access, and waste disposal

The keyword each component listed in table 4,5 and 6. Table 4 is keyword for aspect of product.

Table 4 Keyword Aspect of Products

Component Of Aspect Products	***Keyword***
Dining area and circulation	"cozy","spacious","cramped","organized","clean","dirty","bright","dark","quiet", "spacious", "open".
Provision of food and beverages.	"design","beautiful","modern","traditional","aesthetic","unique", "continental","delicious","delicious","notdelicious","distinctive", "authentic", "special", "special", "original", "savory", "good", "rich", "hot", "cold", "best", "average","appetizing","fresh", "undercooked","overcooked","well prepared","fancy","not fresh".
Supporting facilities.	"colorful", "big", "small", "sufficient", "excessive", "insufficient", "balanced", "satisfactory", "unsatisfactory", "just right", "overpriced", "value for money", "spacious", "crowded", "disorganized", "hot", "cool", "fragrant", "smell", "well maintained", "unmaintained", "complete", "damaged", "available", "unavailable", "easily accessible", "difficult to access", "functional", "clear", "unclear", "informative", "colorful", "monotonous", "safe", "unsafe".

Table 5 Keyword Aspect of Services

Component Aspect of Services	***Keywords***
Standard operating procedures	"organized", "chaotic", "efficient", "inefficient", "routine", "non-routine", "professional", "unprofessional", "standard", "non-standard", "friendly", "less friendly".
Other facilities	"responsive", "indifferent", "kind", "impolite", "swift", "slow", "meticulous", "careless", "fast", "slow", "complicated", "easy", "difficult", "practical", "confusing", "safe", "error-prone", "awake", "unguarded", "poor", "perfect", "inadequate", "clean", "dirty", "hygienic", "unhygienic", "polite", "impolite".

Table 6 Keyword Aspect of Management

Component Aspect Management	***Keywords***
Organization	"structured", "unstructured", "efficient", "chaotic", "neat", "chaotic", "professional", "amateur", "good coordination", "lack of coordination"
Management	"innovative", "stagnant", "progressive", "conservative", "creative", "less creative", "evolving", "stagnating"
Human resources	"attractive", "unattractive", "maintained", "less maintained", "high standard", "poor"
Facilities and infrastructure	"dirty", "sterile", "neglected", "very hygienic", "unhygienic", "guaranteed", "not guaranteed", "strict", "loose", "safe", "risky", "controlled", "vulnerable", "unguarded", "comfortable", "uncomfortable", "conducive", "noisy", "quiet", "crowded", "pleasant", "boring", "harmonious", "environmentally friendly", "not environmentally friendly".

Next, the aspects extract using rules based any key words that described in Table 4, 5 and 6. The algorithm of rules figurated in Figure 1 below.

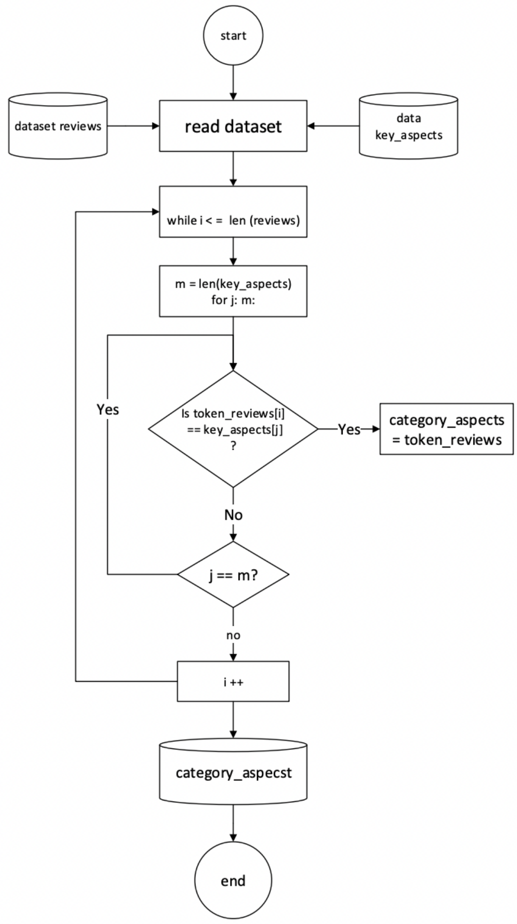

Fig. 2 Flowchart for Extracting Aspects

III. RESULT

A. Sentiment Analysis

Figures 3 shows sentiments of customer of restaurant in Banyumas District. Thre are 3 classes sentiment: positive reach of 85.4%, negative reach of 4.7% and neutral reach of 9.7%.

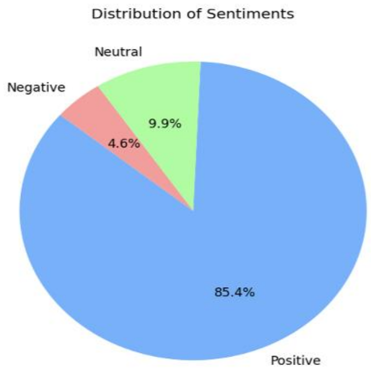

Fig. 3 Sentiment of restaurant in Banyumas District

As in Figure 4, the compound score distribution for positive in range 0.6 to 1.0. This range encompasses the largest number of reviews, peaking at scores around 0.8 to 0.9, each with a frequency exceeding 1,000 reviews.

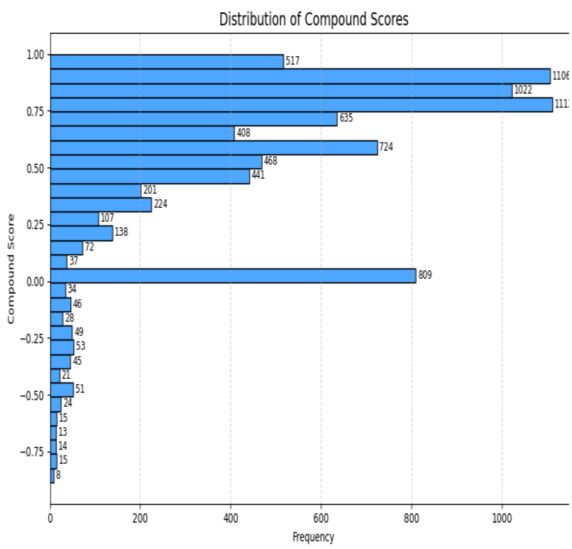

Fig. 4 Compound Score Distribution

Conversely, negative scores are small and evenly distributed in the range below zero. Furthermore, 809 reviews have a neutral score, indicating no particular sentiment bias. This distribution pattern further confirms the previous finding that the majority of reviews are positive

B. Aspect Analysis

Table 7 shows that the product aspect was the most discussed aspect by users, with 7,953 reviews. This indicates that consumers pay close attention to the quality of the food and beverages served by restaurants. Management and services aspects received 4,710 and 5,241 reviews, respectively, indicating that these two aspects also received significant attention. Furthermore, 154 reviews could not be classified into these three aspects due to the limited number of keywords used.

Table 7 Sum of Reviews Each Aspect

Aspects category	**Sum**
Product	7953
Management	4710
Services	5241
No Aspects Detected	154

Dictionary of keyword can be expanded based on further data exploration by adding new words that appear significantly in reviews but were not included in the initial dictionary.

IV. DISCUSSION

As in Table 7, then we analyze keywords which appear most frequently in each category. The keyword most frequently mentioned by customer is shown by word cloud in Figure 5 – 10. In case aspect of product, shown by figure 5 and 6, the biggest word is "Lezat" (Delicious) and "Makanan" (food) - indicating customers love the restaurant because of taste of the food – delicious or not.

Fig. 5 Wordcloud Keyword of Product in Class Positive

Fig. 6 Wordcloud Keyword of Product in Class Negative

Fuhermore, in services case, customers respect to services – shown by wordcloud Figure 7 and 8. Theese shown by words "bagus" (good), "Nyaman" (Comfortable), "ramah" (honest). Although, "ramah" (honest) is not the mosf frequently mentioned, but it appears quite often. But customer also disappointed by services- it shown by the word "melayani" (service), "pelayan" (waiters), "memesan"(order). Theese negative keyword are indicating that the waiter provides poor service especially when the customer makes an order.

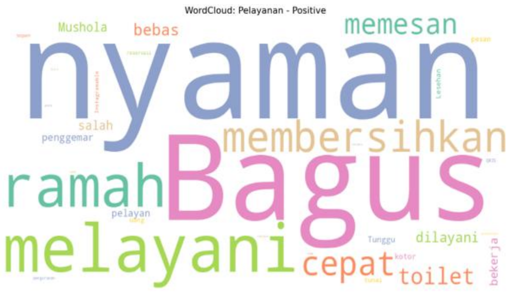

Fig. 7 Wordcloud Keyword of Services In Class Positive

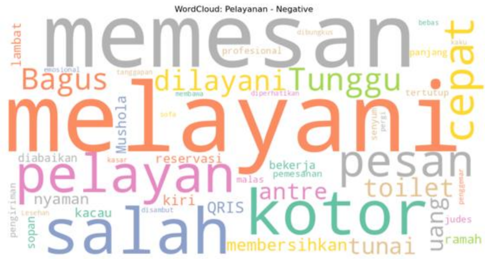

Fig. 8 Wordcloud Keyword of Services in Class Negative

Meanwhile, in terms of management, customers tend to like prices and friendliness- as in Figure 9, keyword “harga” (price), “ramah” (honest) are arising the most. However, they also dislike prices and dirty restaurants- as in Figure 10 shown by keyword “harga” (price) and “kotor” (dirty).

Oddly, the same keyword appears in different sentiment categories—both positive and negative. The keyword "price" evokes negative sentiment in some customers and positive sentiment in others. This could be because each restaurant prioritizes only one or two aspects.This could be a clue for researchers to explore more deeply the factors that influence customer sentiment.

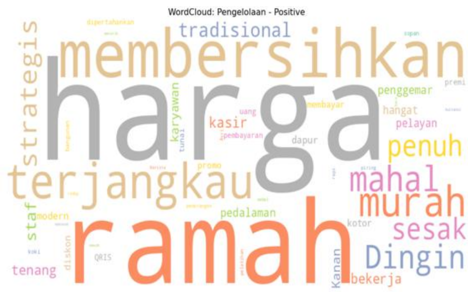

Fig. 9 Wordcloud Keyword of Management In Class Positive

Fig. 10 Wordcloud Keyword of Management In Class Negative

V. CONCLUSSION

The majority restaurant in Banyumas Distriv gain positive sentiment (85.4%), while negative sentiment (4.6%), and neutral sentiment (9.9%). In class positive, customers assign sentiment caused by aspects of product- being the most frequently discussed (7,839 reviews), followed by management (4,608 reviews), and service (4,589 reviews).

Opinion frequently discussed by customer encompass three main aspects. For the product aspect, terms such as "delicious," "food," and "comfortable" were the most frequently mentioned, reflecting that taste, quality, and comfort while enjoying the food are highly valued by visitors. For the service aspect, terms such as "friendly," "serving," and "fast" dominated, indicating that staff attitude and efficiency play a significant role in shaping customer satisfaction. Meanwhile, for the management aspect, topics such as price, cleanliness, and strategic location were frequently discussed, indicating that cleanliness, pricing policies, and restaurant location are important considerations in consumer assessments.

REFERENCES

[1] Menteri Pariwisata Dan Ekonomi Kreatif / Kepala Badan Pariwisata Dan Ekonomi Kreatif Republik Indonesia, *Peraturan Menteri Pariwisata Dan Ekonomi Kreatif / Kepala Badan Pariwisata Dan Ekonomi Kreatif Republik Indonesia Nomor 4 Tahun 2021 Tentang Standar Kegiatan Usaha Pada Penyelenggaraan Perizinan Berusaha Berbasis Risiko Sektor Pariwisata*. 2021.

[2] M. Syamala and N. Nalini, “ABSA: Computational Measurement Analysis Approach for Prognosticated Aspect Extraction System,” *TEM Journal*, vol. 10, no. 1, 2021, doi: 10.18421/TEM101-11.

[3] M. T. Anwar, D. Trisanto, A. Juniar, and F. A. Sase, “Aspect-based Sentiment Analysis on Car Reviews Using SpaCy Dependency Parsing and VADER,” *Advance Sustainable Science, Engineering and Technology*, vol. 5, no. 1, 2023, doi: 10.26877/asset.v5i1.14897.

[4] A. G. Chifu and S. Fournier, “Sentiment Difficulty in Aspect-Based Sentiment Analysis,” *Mathematics*, vol. 11, no. 22, 2023, doi: 10.3390/math11224647.

[5] W. Parasati, F. Abdurrachman Bachtiar, and N. Y. Setiawan, “Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naïve Bayes Classifier,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 4, 2020.

[6] A. Nofandi, N. Y. Setiawan, and D. W. Brata, “Analisis sentimen ulasan pelanggan dengan Metode Support Vector Machine (SVM) untuk peningkatan kualitas layanan pada Restoran Warung Wareg,” *Jurnal Pengembangan*

Teknologi Informasi dan Ilmu Komputer, vol. 7, no. 1, 2023.

[7] V. Fadillah[1], F. Hamami[2], and R. Andreswari[3], "Analisis Sentimen Berbasis Aspek Terhadap Ulasan Pengguna Aplikasi Pegadaian Digital Dengan Multiclass Multioutput Menggunakan Algoritma Support Vector Machine," *Kesatria : Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 4, no. 4, 2023.

[8] G. Radiena and A. Nugroho, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Kai Access Menggunakan Metode Support Vector Machine," *Jurnal Pendidikan Teknologi Informasi (JUKANTI)*, vol. 6, no. 1, 2023, doi: 10.37792/jukanti.v6i1.836.

[9] B. Shin, S. Ryu, Y. Kim, and D. Kim, "Analysis on Review Data of Restaurants in Google Maps through Text Mining: Focusing on Sentiment Analysis," *Journal of Multimedia Information System*, vol. 9, no. 1, 2022, doi: 10.33851/jmis.2022.9.1.61.

[10] D. Arianto and I. Budi, "Analisis Sentimen Berbasis Aspek dan Pemodelan Topik pada Candi Borobudur dan Candi Prambanan," *MULTINETICS*, vol. 8, no. 2, 2023, doi: 10.32722/multinetics.v8i2.5056.

[11] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, 2021, doi: 10.30865/mib.v5i2.2845.

[12] F. Illia, M. P. Eugenia, and S. A. Rutba, "Sentiment Analysis on PeduliLindungi Application Using TextBlob and VADER Library," *Proceedings of The International Conference on Data Science and Official Statistics*, vol. 2021, no. 1, 2022, doi: 10.34123/icdsos.v2021i1.236.

[13] K. Barik and S. Misra, "Analysis of customer reviews with an improved Vader lexicon classifier," *J Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-023-00861-x.

[14] B. M. Alenzi, M. B. Khan, M. H. A. Hasanat, A. K. J. Saudagar, M. Alkhathami, and A. Altameem, "Automatic Annotation Performance of TextBlob and VADER on Covid Vaccination Dataset," *Intelligent Automation and Soft Computing*, vol. 34, no. 2, 2022, doi: 10.32604/iasc.2022.025861.

[15] B. Liu *et al.*, "Noise Learning for Text Classification: A Benchmark," in *Proceedings - International Conference on Computational Linguistics, COLING*, 2022.

[16] X. Chen, B. He, K. Hui, L. Sun, and Y. Sun, "Dealing with textual noise for robust and effective BERT re-ranking," *Inf Process Manag*, vol. 60, no. 1, 2023, doi: 10.1016/j.ipm.2022.103135.

[17] L. V. Subramaniam, S. Roy, T. A. Faruquie, and S. Negi, "A survey of types of text noise and techniques to handle noisy text," in *ACM International Conference Proceeding Series*, 2009. doi: 10.1145/1568296.1568315.

[18] M. Işik and H. Dağ, "The impact of text preprocessing on the prediction of review ratings," 2020. doi: 10.3906/elk-1907-46.

[19] L. Ardiani, H. Sujaini, and T. Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 2, 2020, doi: 10.26418/justin.v8i2.36776.

[20] E. Elinda, H. Yuliansyah, and M. I. A. Latiffi, "Sentiment Analysis of the Sheikh Zayed Grand Mosque's Visitor Reviews on Google Maps Using the VADER Method," *International Journal of Advances in Data and Information Systems*, vol. 5, no. 1, pp. 71–84, Apr. 2024, doi: 10.59395/ijadis.v5i1.1320.