Journal of Emerging Information Systems and Business Intelligence
ISSN: 2774-3993
Journal homepage: https://ejournal.unesa.ac.id/index.php/JEISBI/

Importance Performance Analysis (IPA) of Google Reviews Sentiments Based on SERVQUAL Dimension for Public Health Center Service in Surabaya

Octania Sriwahyuni[1], Wiyli Yustanti[2]

[1, 2]*Universitas Negeri Surabaya, Surabaya, Indonesia*

Octania.22128@mhs.unesa.ac.id, *wiyliyustanti@unesa.ac.id*

ABSTRACT

Google Reviews can serve as a digital mirror to gauge how the public evaluates the services of community health centers. This study focuses on analyzing the sentiment of Google Reviews for community health centers in Surabaya, mapping reviews into the SERVQUAL dimensions using Gap Analysis and Importance–Performance Analysis (IPA), and identifying the most influential keywords via TF-IDF within a GUI system. This study applies the Knowledge Discovery in Databases (KDD) workflow. Data was obtained by scraping reviews from 63 community health centers in Surabaya. Subsequently, sentiment was determined based on user ratings, then classified into the five SERVQUAL dimensions, and analyzed using the GAP analysis and Importance-Performance Analysis (IPA). The results indicate that positive public perceptions predominate. However, all dimensions still show negative scores, suggesting that service quality has not yet fully met user expectations. In the IPA analysis, Responsiveness, Assurance, and Empathy are categorized in Quadrant II as aspects that require maintenance, while Tangibles and Reliability fall into Quadrant III as low-priority aspects. Notably, no dimension is in Quadrant I. Additionally, TF-IDF successfully captures keywords such as “queue,” “long,” “friendly,” “clean,” and “procedure,” and has been successfully implemented in the GUI for automatic classification. Building on these results, this study confirms that digital reviews combined with sentiment analysis, SERVQUAL, and Importance-Performance Analysis (IPA) can serve as a more objective, practical, and sustainable evaluation tool for Public health center services.

Keyword: Sentiment Analysis, SERVQUAL, IPA, TF-IDF, Google Reviews, Public health center.

Article Info:
Article history:
Received May 20, 2026
Revised June 03, 2026
Accepted June 10, 2026

Corresponding Author
Octania Sriwahyuni
Universitas Negeri Surabaya, Surabaya, Indonesia
Octania.22128@mhs.unesa.ac.id

1. INTRODUCTION

As frontline healthcare facilities, community Public health centers play a major role in providing healthcare services to the public [1]. The quality of these services is a key determinant of patient satisfaction [2]. Therefore, service quality must be continuously monitored to ensure that the quality of care remains consistent and continues to improve [3]. However, traditional evaluation methods such as surveys and interviews remain suboptimal due to their time-consuming nature, high costs, and limited respondent reach [4]. In the digital age, people now frequently share their experiences and evaluations online, such as through

Google Reviews [5]. These reviews represent users' impressions of the services they have experienced, making them a valuable source of information for assessing service quality in a broader and more up-to-date manner [6]. With the abundance of review data, sentiment analysis can help identify public opinion trends regarding Public health center services in a more practical way [7].

Nevertheless, much of the existing sentiment analysis research still limits itself to categorizing opinions as positive, negative, or neutral, without examining their relationship to service quality dimensions in greater detail [8]. Therefore, a more comprehensive method is needed that incorporates the SERVQUAL model, which consists of tangibles, reliability, responsiveness, assurance, and empathy [9]. Furthermore, Gap Analysis is used to identify the gap between service expectations and actual performance, while Importance-Performance Analysis (IPA) is used to determine the aspects most in need of improvement based on importance scores and service performance [10]. This study processes Google Reviews from 63 Public Health Centers in Surabaya, collected via web scraping and organized using Knowledge Discovery in Databases (KDD) to ensure data is collected and analyzed systematically [11]. User sentiment is determined through ratings, then combined with the SERVQUAL dimensions for further analysis using GAP and IPA [12]. The TF-IDF method is also used to identify dominant terms in the comments as an indication of public assessment of service quality [13].

The objective of this study is to examine public perceptions of Public Health Center services based on Google Reviews, while providing a more measurable evaluation through the integration of sentiment analysis, SERVQUAL, GAP, Importance-Performance Analysis (IPA), and TF-IDF [14]. The results are expected to support continuous improvements in Public Health Center services [15].

2. METHODS

This study employs a descriptive quantitative approach using the Knowledge Discovery in Databases (KDD) framework to identify patterns in user reviews [16]. The data was collected from Google Reviews via web scraping using Octoparse for 63 community health centers in Surabaya, yielding approximately 25,000 reviews containing text comments and rating scores [17]. The research workflow consists of data collection, preprocessing, transformation, and analysis. During the preprocessing stage, the data undergoes cleaning, standardization, tokenization, removal of irrelevant words, and stemming to ensure the text is clean and easier to process [18]. User sentiment was analyzed based on ratings, then categorized into positive, neutral, and negative [19]. Subsequently, the reviews were linked to the five SERVQUAL dimensions—tangibles, reliability, responsiveness, assurance, and empathy using keyword matching [20]. The mapped data is then encoded into numbers using MultiLabel Binarizer for quantitative analysis [21]. The TF-IDF method is used to convert text into numerical values while highlighting the most frequently occurring words [22]. Random Forest is also utilized as an automatic classification model and for performance evaluation [23]. Service quality is measured using GAP analysis with the formula.

$$\textbf{GAP = Importance – Performance} \quad (1)$$

Followed by the calculation of the Conformity Index (CI) using the formula:

$$\boldsymbol{Ci} = \frac{\boldsymbol{Xi}}{\boldsymbol{Yi}} \times \mathbf{100}\% \quad (2)$$

The Importance-Performance Analysis (IPA) method is used to determine service improvement priorities by mapping the average importance and performance into four quadrants [24].

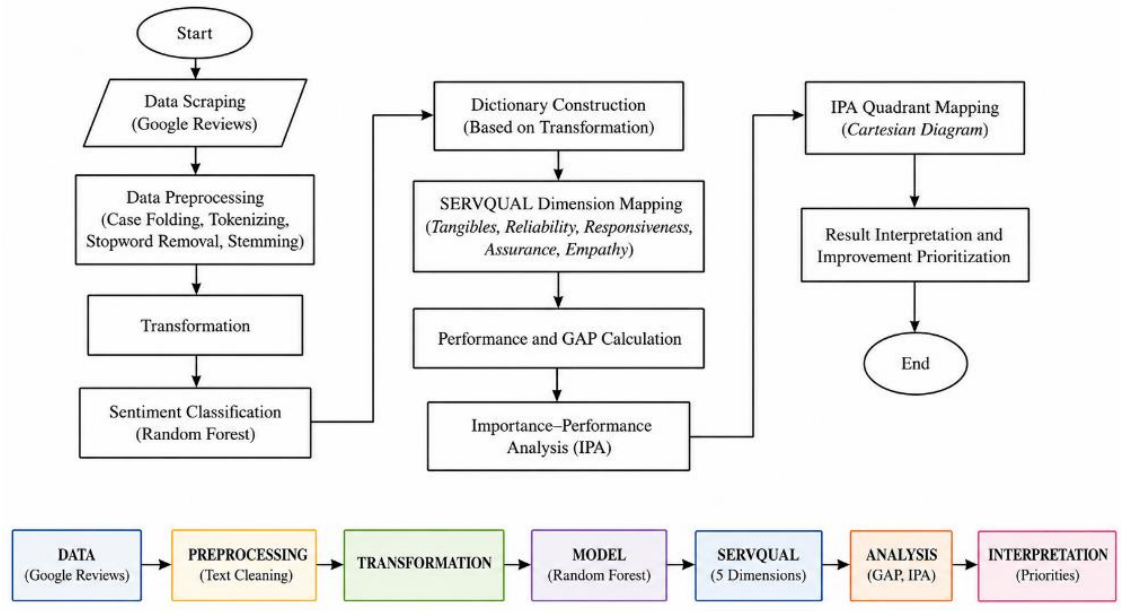


Figure 1 Knowledge Data Discovery

3. RESULTS AND DISCUSSION

This section presents the results of the processing and analysis of Google Reviews for Public Health Centers (Puskesmas) in Surabaya. The analysis was conducted to assess user sentiment based on ratings, categorize reviews into the five dimensions of SERVQUAL, and calculate the GAP and CI between performance and expectations. The results were then analyzed using the IPA method to determine priorities for service improvement.

3.1 Data Selection

Research data was collected by tracking user reviews on Google Reviews using Octoparse. The data source consisted of 63 Public health centers in Surabaya, including public comments and service ratings on a scale of one to five stars. The data collection process yielded approximately 25,000 reviews. Next, the data was cleaned by eliminating incomplete records, such as reviews without text or star ratings, to ensure the data remained valid and ready for use. The final dataset contained the names of the Public Health Center, user identities, review dates, ratings, and comment content, which were then compiled into a single dataset for the next stage of analysis.

Table 1. Data Variable

Column Name	Data Type	Description
Review_Id	Float64	The unique ID for each user review on Google Reviews.
Review_Text	Object	The review submitted by the user regarding the service provided.
Ratings	Int64	A rating on a scale of 1 to 5 given by the user
Username	Object	The name of the user who wrote the review.
Publish At Date	Object	The date the review was published on Google Reviews

Table 1 presents the variables collected from Google Reviews, including review ID, review text, ratings, username, and publication date. These variables serve as the primary dataset for analyzing user perceptions of public health center services in Surabaya.

3.2 Preprocessing

The data obtained from web scraping is still unstructured and therefore requires preprocessing. This process includes case folding, tokenization, stopword removal, and stemming to clean and organize the text data. The end result is text that is more organized, cleaner, and ready for further analysis. Table 2 shows examples of review text before and after the preprocessing stage. This process successfully transformed unstructured review data into cleaner and more standardized text, making it suitable for further sentiment analysis and SERVQUAL dimension mapping.

Table 2. Result Text Review

Text	Text_Final
puskesmas jagir surabaya top markotop, pelayanannya baik dan ramah, tempatnya bersih dan asri, sukses terus puskesmas jagir 👍👍👍👍👍	puskesmas jagir surabaya top markotop pelayanannya baik dan ramah tempatnya bersih dan asri sukses terus puskesmas jagir
Bagian pelayanan informasinya sangat buruk, enggan memberikan arahan saat pendaftaran. Mohon dievaluasi!!	bagian pelayanan informasinya sangat buruk enggan memberikan arahan saat pendaftaran mohon dievaluasi
Pelayanan cepat, perawat nya baik apalagi mahasiswa PKL stikes wb	pelayanan cepat perawat nya baik apalagi mahasiswa pkl stikes wb
Parah...pokoknya Parah...tanya baik baik dijawab judes...tau semua orang capek tapi ya jawab biasa aja..apalagi ini ugd...males kesini sebenarnya barusan tapi urgensi....dulu nya dah kecewa Sekarang tambah kecewa..mana penerapan 5S nya ibu" dan bapak bapak	parah pokoknya parah tanya baik baik dijawab judes tau semua orang capek tapi ya jawab biasa aja apalagi ini ugd males kesini sebenarnya barusan tapi urgensi dulu nya dah kecewa sekarang tambah kecewa mana penerapan s nya ibu dan bapak bapak
untuk pelayanan pemeriksaan hamil memuaskan dan pelayanannya cepat bidan bidannya juga ramah dan memberikan penjelasan yang jelas	pelayanan pemeriksaan hamil memuaskan pelayanannya cepat bidan bidannya ramah penjelasan

3.3 Transformation

Data transformation is performed by converting user ratings into negative, neutral, and positive sentiment labels, and then converting them into numerical values for classification and further analysis.

Table 3. Transformation Result

Text_Final	Ratings	Sentiment	Sentiment_label
puskesmas jagir surabaya top markotop pelayanannya ramah tempatnya bersih asri sukses puskesmas jagir	5	3	Positive
pelayanan informasinya buruk enggan arahan pendaftaran mohon dievaluasi	1	1	Negative
pelayanan cepat perawat nya mahasiswa pkl stikes wb	5	3	Positive
pelayanan diping pong	1	1	Negative
kasir suruh antre nunggu lama ga panggil panggil dokternya main hp negur bilang panggil belum inisial dokter r a p	1	1	Negative

Based on Transformation, reviews with low ratings were classified as negative sentiment, those with moderate ratings as neutral, and those with high ratings as positive. This classification was used because ratings reflect user satisfaction with the service. Subsequently,

each sentiment was converted into a numerical value to facilitate further data processing and analysis.

3.4 Analysis Sentimen

The next step is to analyze the sentiment of Google Reviews for community health centers in Surabaya. Sentiment categories were determined based on ratings: 1–2 for negative, 3 for neutral, and 4–5 for positive. The results are represented by the `sentiment_label` attribute, which reveals the public's overall assessment of the quality of services provided by community health centers.

Table 4 Result Analysis Sentiment

Sentiment Category	Count Distribution	Percentage (%)
Positive	14458	70.66
Neutral	5448	26.63
Negative	544	2.71

Table 4 shows that positive sentiment dominates the review results, accounting for 70.66%. This suggests that the public generally appreciates the services provided by community health centers in Surabaya. Neutral sentiment accounts for 26.63% of the reviews, while negative sentiment represents only 2.71%. Figure 2 further illustrates the distribution of sentiment categories.

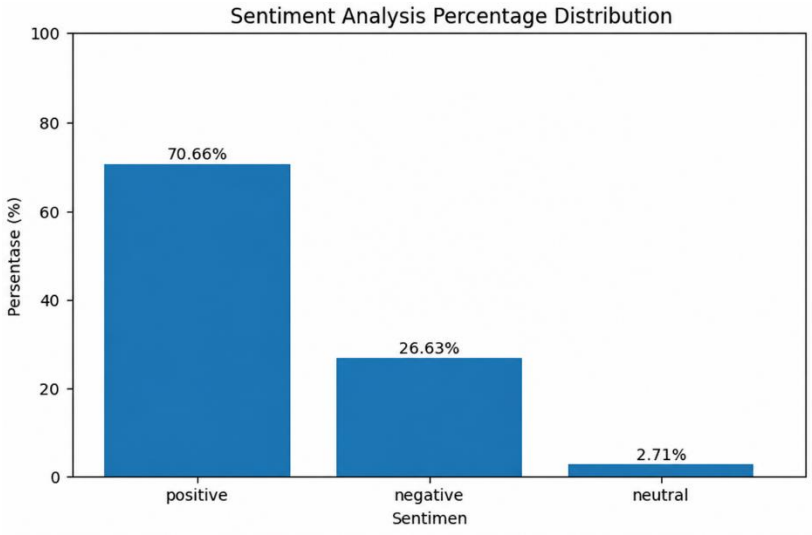


Figure 2. Sentiment Distribution

Figure 2 shows that positive sentiment is the most dominant category, followed by neutral and negative sentiment. These findings indicate that community health center services in the Surabaya metropolitan area generally provide a fairly positive experience for residents. The results of this sentiment analysis were then used as a basis for mapping the Servqual dimensions and conducting an Importance-Performance Analysis (IPA).

3.5 Service Quality (SERVQUAL)

An analysis of the service quality at Public Health centers in Surabaya was conducted using Google Reviews and the SERVQUAL and Importance Performacne Analysis methods. The processed reviews were classified into the five SERVQUAL dimension Tangibles, Reliability, Responsiveness, Assurance, and Empathy, based on keywords. Each review could be assigned to more than one dimension if it contained words reflecting multiple aspects of service.

Table 5. Distribution of Counts by Dimension

Dimension	Count Distribution
Tangibles	4342
Responsiveness	14749
Assurance	13231
Empathy	8276
Reliability	4239

Table 5 shows that the responsiveness dimension was the most frequently mentioned aspect among all SERVQUAL dimensions. This finding indicates that users placed greater emphasis on service speed, staff responsiveness, and how service requests were handled. On the other hand, the dimensions of assurance and empathy also had a notably high frequency of mention. This means that the expertise of medical staff, their professionalism, friendly attitude, and attentiveness also shape how the public assesses service quality.

The data is converted into numerical form using the MultiLabelBinarizer and TF-IDF. This technique is used to describe the dimensions of SERVQUAL and to determine the weight of each word in every review. From these results, we can identify the key words that reflect the public's views on the quality of service at the Public Health center. Table 6 presents the dominant keywords identified in each SERVQUAL dimension.

Table 6. Results of Dominant Words by Dimension

Dimension	Dominant Words
Tangibles	'bersih','kotor','ruang','ruangan','fasilitas','toilet','parkir','peralatan','ac','pendingin','nyaman','sempit','luas','rapi','kumuh','lengkap'
Reliabilty	'tepat','telat','sesuai','akura,'konsisten','janji','hasil','pelayanan
Responsiveness	'cepat','lambat','lama','antrian','tanggap','respon','menunggu', 'datang','dipanggil','proses','pelayanan cepat'
Assurance	aman,'tidak aman', 'percaya', 'yakin','profesional, 'kompeten', berpengalaman, jelas','penjelasan', meyakinkan', dokter', 'petugas','staf'
Empathy	'ramah','jutek','cuek','peduli', 'perhatian', 'sabar','kasar', 'membantu','dibantu','melayani', 'dilayani','sopan','tidak sopan', 'komunikatif'

Based on the results at table 6, each dimension has a dominant word pattern that reflects the focus of service evaluation. The frequency of these words indicates that the public places significant emphasis on facilities, speed of service, the competence of medical staff, and the friendliness of personnel. These four elements form a crucial foundation for evaluating the quality of community health center services.

Table 7. Result GAP & Conformity Index

Dimension	GAP	Percentage Change (%)
Tangibles	-2.65	46.96
Responsiveness	-2.54	49.19
Assurance	-2.48	50.30
Empathy	-2.34	53.11
Reliability	-2.68	47.41

Subsequently, GAP analysis and the Conformity Index (CI) were used to assess the gap between service performance and user expectations. The GAP was calculated by subtracting the performance score from the importance score, while the CI indicates the level

of service conformity as a percentage. Details of the calculation results are presented in Table 7.

The assessment results show that all SERVQUAL dimensions still exhibit a negative GAP, with CI scores below 100%. This indicates that the services provided have not yet fully met public expectations. The Assurance dimension appears to be the closest to meeting public expectations, while Tangibles and Empathy are the aspects with the largest quality gaps and therefore require prioritization for improvement.

3.6 Analysis Importance-Performance Analysis (IPA)

After the performance and importance scores have been calculated, the next step focuse on mapping. The Importance–Performance Analysis (IPA) to identify service aspects that need to be prioritized for development within each SERVQUAL dimension. The performance score reflects the quality of service as perceived by users, while The importance score indicates the level of urgency for each dimension based on the frequency of its mention in reviews.

Table 8. Result Importance-Performance Analysis

Dimension	Distrubution	Performance	Importance
Tangibles	4342	0.47	0.09
Responsiveness	14749	0.50	0.28
Assurance	8276	0.53	0.18
Empathy	4239	0.47	0.09
Reliability	13231	0.49	0.31

Table 8 shows that the assurance dimension achieved the highest performance score of 0.53, followed by responsiveness at 0.50. This indicates that service quality in these two dimensions was rated higher than in other aspects. On the other hand, Reliability, Tangibles, and Empathy achieved slightly lower performance scores. In terms of importance, users highlighted the dimensions of Reliability (0.31) and Responsiveness (0.28) as the primary factors in their service evaluations. The results were then mapped into an Importance–Performance Analysis (IPA) diagram to identify service improvement priorities across each SERVQUAL dimension.

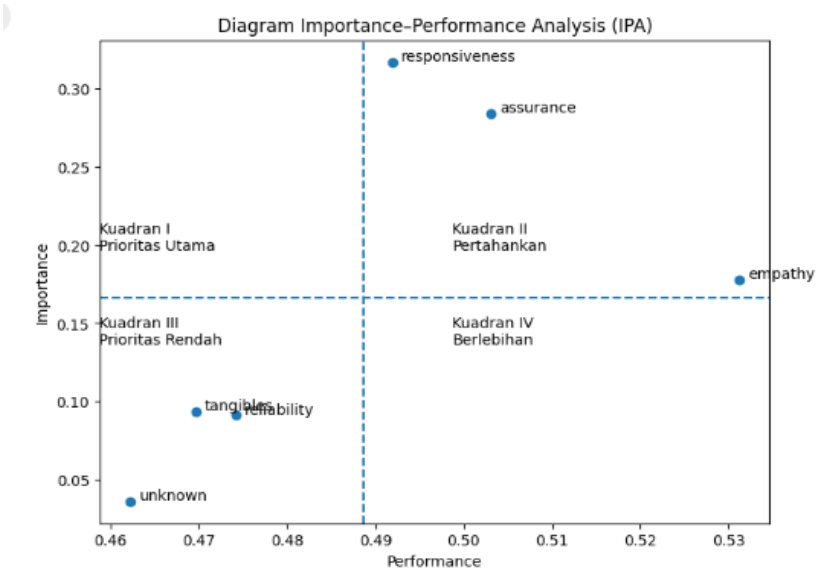


Figure 3. Diagram Importance - Performance Analysis

Figure 3 illustrates the Importance–Performance Analysis (IPA) mapping of all SERVQUAL dimensions. The responsiveness and assurance dimensions are positioned in Quadrant II, indicating high importance and satisfactory performance that should be maintained. The empathy dimension is located in Quadrant IV, suggesting that its performance exceeds its perceived importance. Meanwhile, the tangibles and reliability dimensions are positioned in Quadrant III, indicating relatively lower importance and performance, and are therefore considered lower priorities for immediate improvement.

Table 9. Result Random Forest

Dimension	Accuracy	Precision	Recall	F1-Score	Support
Tangibles	0.965	0.99	0.84	0.91	874
Reliablity	0.944	0.96	0.77	0.86	889
Responsiveness	0.988	1.00	0.99	0.99	2969
Assurance	0.987	1.00	0.97	0.98	1631
Empathy	0.987	0.98	1.00	0.99	2461

Table 9 presents the performance evaluation results of the Random Forest model across all SERVQUAL dimensions. The model achieved excellent classification performance, with accuracy ranging from 0.944 to 0.988 and F1-scores between 0.86 and 0.99. These findings indicate that the model is capable of accurately identifying service categories and is suitable for implementation in an automated service quality classification system.

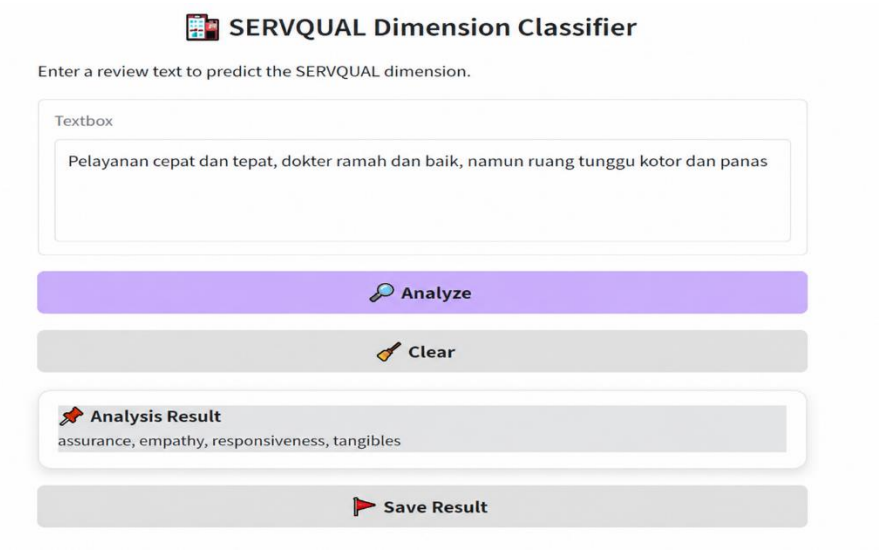


Figure 4. Interface Dimension Classifier

Figure 4 presents the interface of the developed SERVQUAL Dimension Classifier, which operates through an interactive GUI to automatically identify SERVQUAL dimensions from customer review text. During testing, the system successfully mapped reviews into the relevant service categories, including Responsiveness, Assurance, Empathy, and Tangibles, indicating that the multi-label classification performed effectively. In addition, all GUI features functioned properly in processing input, executing analysis, and displaying results, demonstrating the practical applicability of the system for automated service quality evaluation.

CONCLUSION

This study successfully demonstrated that sentiment analysis of Google Reviews can be used as an effective approach to evaluate the service quality of public health centers (Puskesmas) in Surabaya through the integration of SERVQUAL, GAP Analysis, Importance–Performance Analysis (IPA), TF-IDF, and Random Forest classification. The findings indicate that positive sentiment dominates public perceptions; however, all SERVQUAL dimensions still show negative GAP values, suggesting that service performance has not yet fully met user expectations. The Importance-Performance Analysis results revealed that responsiveness and assurance are positioned in Quadrant II and should be maintained, while tangibles and reliability are categorized as lower-priority aspects for improvement in Quadrant III. Meanwhile, empathy falls into Quadrant IV, indicating performance that exceeds its perceived importance. Additionally, TF-IDF successfully identified dominant keywords representing users' service experiences, and the Random Forest model achieved high classification performance across all SERVQUAL dimensions. The implementation of the model into a GUI-based system further demonstrates its practical applicability for automatic service quality evaluation. Overall, this study confirms that digital review analysis can serve as an objective, efficient, and sustainable tool to support continuous improvement in public health center services.

REFERENCES

[1] Saputra, A., Hajar, S., & Sari, M. T. (2024). Analisis kebijakan kesehatan dalam meningkatkan mutu pelayanan kesehatan Puskesmas di Kota Medan. Kebijakan: Jurnal Ilmu Administrasi, 15(2). https://doi.org/10.23969/kebijakan.v15i02.10182

[2] Mulyani, A., Fudholi, A., & Satibi. (2021). Analisis tingkat kepuasan pasien terhadap pelayanan kefarmasian menggunakan model SERVQUAL di Puskesmas Kabupaten Garut. Majalah Farmaseutik, 17(3), 284–295. https://doi.org/10.22146/farmaseutik.v17i3.54017

[3] Masriyatun, Wahyudin, N., & Yusnita, M. (2022). Kualitas pelayanan Puskesmas Lubuk Besar dengan pendekatan Importance Performance Analysis (IPA) dan Customer Satisfaction Index (CSI). https://doi.org/10.17358/jabm.8.3.891

[4] Ayudia, S., Nadeak, B., & Suyaman, D. J. (2021). Evaluasi mutu pelayanan Puskesmas terakreditasi berdasarkan indeks kepuasan masyarakat di Kabupaten Karawang. Syntax Literate: Jurnal Ilmiah Indonesia, 6(6), 3037–3048. https://doi.org/10.36418/syntax-literate.v6i6.2340

[5] Rahim, A. I. A., Ibrahim, M. I., Chua, S.-L., & Musa, K. I. (2021). Hospital Facebook reviews analysis using a machine learning sentiment analyzer and quality classifier. Healthcare, 9(12), 1679. https://doi.org/10.3390/healthcare9121679

[6] Song, B., Lee, C., Yoon, B., & Park, Y. (2016). Diagnosing service quality using customer reviews: An index approach based on sentiment and gap analyses. Service Business, 10(4), 775–798. https://doi.org/10.1007/s11628-015-0290-1

[7] Wulandari, R., & Hidayanto, A. N. (2024). Measuring contact tracing service quality using sentiment analysis: A case study of PeduliLindungi Indonesia. Quality & Quantity, 58(2), 1409–1424. https://doi.org/10.1007/s11135-023-01695-8

[8] Zhou, S., Xie, K., & Zhang, X. (2025). Identify doctors' online performance and prioritize service attributes: A framework with aspect-based sentiment analysis and empirical investigation. Digital Health. https://doi.org/10.1177/20552076251353320

[9] Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing, 64(1),

12–40. https://doi.org/10.2307/1251430

[10] Hermawan, A., Jowensen, I., Junaedi, & Edy. (2023). Implementasi text-mining untuk analisis sentimen pada Twitter dengan algoritma Support Vector Machine. Jurnal Sains dan Teknologi, 12(1), 129–137. https://doi.org/10.23887/jstundiksha.v12i1.52358

[11] Rahim, A. I. A., Ibrahim, M. I., Chua, S.-L., & Musa, K. I. (2021). Hospital Facebook reviews analysis using a machine learning sentiment analyzer and quality classifier. Healthcare, 9(12), 1679. https://doi.org/10.3390/healthcare9121679

[12] Ramdani, R., Dewi, P. S., Islamiyah, A. N., Hermanto, F., & Putri, I. R. P. (2026). Analisis tingkat kepuasan pasien JKN terhadap pelayanan kefarmasian menggunakan metode SERVQUAL dan Importance Performance Analysis (IPA) di salah satu Puskesmas Kota Cimahi. Pharmacoscript, 9(1), 115–122. https://doi.org/10.36423/pharmacoscript.v9i1.2348

[13] Hermawan, A., Jowensen, I., Junaedi, & Edy. (2023). Implementasi text-mining untuk analisis sentimen pada Twitter dengan algoritma Support Vector Machine. Jurnal Sains dan Teknologi, 12(1), 129–137. https://doi.org/10.23887/jstundiksha.v12i1.52358

[14] Gopal, R. K., & Bhargavi, R. (2019). Performance evaluation of best feature subsets for sentiment analysis using machine learning algorithms. Applied Artificial Intelligence. https://doi.org/10.1080/08839514.2019.1592343

[15] Monica, D., Irianto, & Afandi, A. M. (2023). SERVQUAL method analysis for general patient health service satisfaction level. Jurnal Ipteks Terapan, 17(1), 23–30. https://doi.org/10.22216/jit.v17i1.1482

[16] Panggabean, H. L., Gultom, F., & Nasution, M. B. K. (2025). Analisis sentimen terhadap Rumah Sakit Tarutung berdasarkan Google Maps Review menggunakan algoritma Naïve Bayes dan pembobotan kata TF-IDF. Jurnal Pendidikan Sains dan Komputer, 5(2). https://doi.org/10.47709/jpsk.v5i02.7352

[17] Ikhsan, I. N. T., Rudiman, R., & Fendy, Y. (2024). Analisis sentimen pada ulasan Google Maps terhadap pelayanan BPJS Kesehatan menggunakan TF-IDF dan KNN. Jurnal Teknologi Informasi, 18(2). https://doi.org/10.47111/jti.v18i2.15010