e-ISSN 2540-9840 p-ISSN 1412-3770 Jurnal Ilmiah Sains, April 2026, 26(1): 86-104
Accredited by Ministry of Education, Culture, Research and Technology No: 105/E/KPT/2022
DOI: https://doi.org/10.35799/jis.v26i1.67193
Available online at https://ejournal.unsrat.ac.id/v3/index.php/jis

Sentiment Analysis of Healthcare Services at RSUD Soe Using Machine Learning and Latent Dirichlet Allocation

Agatha Marilin Saekoko[1*], Hindriyanto Dwi Purnomo[1)], Yessica Nataliani[1)]

[1,2,3)] **Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia**
*Corresponding author: saekokoagatha21@gmail.com

ABSTRACT

Healthcare services constitute a crucial aspect in improving public well-being. Every individual has the right to receive healthcare services that are of high quality, safe, efficient, and affordable. This study aims to identify and analyze public perceptions and sentiments toward healthcare services at RSUD Soe, as well as to evaluate the performance of several machine learning methods in classifying such sentiments. The data were collected from 278 respondents through a Likert-scale questionnaire that represents perceptions and levels of satisfaction regarding various service aspects. Sentiment analysis was conducted using four machine learning algorithms, namely Naïve Bayes, C4.5, Random Forest, and Support Vector Machine. The results indicate that Naïve Bayes achieved the highest accuracy of 82.14 percent, followed by SVM at 80 percent, Random Forest at 79 percent, and C4.5 at 73.21 percent. This study also applied the Latent Dirichlet Allocation (LDA) method to identify the main themes within public feedback. LDA generated twelve topics reflecting key issues such as waiting time, availability of medical personnel, facility cleanliness, and the attitudes of healthcare staff. The majority of comments exhibited positive sentiment, particularly concerning staff friendliness and service quality. These findings were used to formulate improvement recommendations, including enhancing service quality, increasing the number of medical personnel, and optimizing facilities. This research demonstrates that a data-driven quantitative approach is effective in evaluating healthcare service quality and supporting more targeted decision-making. The results are expected to assist RSUD Soe in continuously and effectively improving service quality.

Keywords: Healthcare Services; Latent Dirichlet Allocation; Machine Learning; RSUD Soe; Sentiment Analysis

Analisis Sentimen Layanan Kesehatan RSUD Soe Menggunakan Machine Learning dan Latent Dirichlet Allocation

ABSTRAK

Pelayanan kesehatan merupakan aspek penting dalam meningkatkan kesejahteraan masyarakat. Setiap individu berhak memperoleh layanan yang berkualitas, aman, efisien, dan terjangkau. Penelitian ini bertujuan untuk mengidentifikasi dan menganalisis persepsi dan sentimen publik terhadap layanan kesehatan di RSUD Soe, serta mengevaluasi kinerja beberapa metode pembelajaran mesin dalam mengklasifikasikan sentimen tersebut.. Data dikumpulkan dari 278 responden melalui kuesioner skala Likert yang menggambarkan persepsi dan tingkat kepuasan terhadap berbagai aspek layanan. Analisis sentimen dilakukan menggunakan empat algoritma machine learning, yaitu Naïve Bayes, C4.5, Random Forest, dan Support Vector Machine. Hasil menunjukkan bahwa Naïve Bayes memiliki akurasi tertinggi sebesar 82,14%, disusul SVM dengan 80%, Random Forest dengan 79%, dan C4.5 dengan 73,21%. Penelitian ini juga menerapkan metode Latent Dirichlet Allocation (LDA) untuk mengidentifikasi tema utama dalam masukan masyarakat. LDA menghasilkan dua belas topik yang mencerminkan isu penting seperti waktu tunggu, ketersediaan tenaga medis, kebersihan fasilitas, dan sikap petugas kesehatan. Mayoritas komentar menunjukkan sentimen positif, terutama terkait keramahan staf dan kualitas pelayanan. Temuan ini digunakan untuk merumuskan rekomendasi perbaikan, termasuk peningkatan kualitas layanan, penambahan tenaga medis, dan optimalisasi fasilitas. Penelitian ini menunjukkan bahwa pendekatan kuantitatif berbasis data efektif dalam mengevaluasi kualitas layanan kesehatan dan mendukung pengambilan keputusan yang lebih tepat sasaran. Hasil penelitian ini diharapkan dapat membantu RSUD Soe meningkatkan mutu layanan secara berkelanjutan dan efektif.

Kata kunci: Analisis sentimen; latent dirichlet allocation; machine learning; pelayanan kesehatan; RSUD Soe

(Article History: Received 01-04-2026; Accepted 15-04-2026; Published 22-04-2026)

INTRODUCTION

Healthcare services are a critical factor in improving the level of well-being of individuals worldwide. Every person has the right to obtain healthcare services that are high-quality, safe, efficient, and affordable, as mandated by Article 19 of Law No. 36 of 2009 (Republic of Indonesia, 2009). Public sentiment analysis toward healthcare services serves as a key indicator reflecting the quality of services provided by a healthcare facility, as patient feedback has been widely utilized to evaluate healthcare performance (Greaves *et al*., 2013; Sarker *et al*., 2015).

Sentiment analysis has a significant impact across various fields, one of which is healthcare services that benefit from technological advancements. Sentiment analysis is an emerging trend that helps healthcare service providers gain a competitive advantage in understanding and improving patient experiences (Lai & Mafas, 2022). Many patients use the internet to provide reviews related to healthcare services, and these online expressions can be analyzed to understand patient experiences and concerns more effectively (Abd-alrazaq *et al.*, 2020). Patient opinions or reviews are important for healthcare providers in improving the quality of their services. These reviews provide valuable insights into healthcare services and assist new patients in making appropriate decisions regarding health, treatment, or the selection of healthcare service providers.

Previous studies have applied various machine learning techniques and topic modeling approaches to analyze healthcare-related text data (Paul & Dredze, 2014; Sharma & Dey, 2012). For example, a study (Rao *et al*., 2024) applied Latent Dirichlet Allocation (LDA) and demonstrated that supervised learning methods produced better classification accuracy compared to unsupervised approaches. A Study by (Mohd Sofi & Selamat, 2023) combined LDA with Term Frequency–Inverse Document Frequency (TF-IDF) and compared the performance of Support Vector Machine (SVM) and Naïve Bayes algorithms, where SVM achieved an accuracy of up to 85%. This research indicates that aspect-based sentiment analysis can enhance public understanding of health-related issues such as COVID-19. Several studies (Lai & Mafas, 2022) also emphasize linguistic challenges such as negation, irony, and coreference in healthcare sentiment analysis, which affect classification accuracy.

Recent studies have shown that advanced machine learning and deep learning models are effective in handling complex linguistic patterns in sentiment analysis (Shickel *et al*., 2018). Meanwhile, research conducted at RSUD Kota Cilegon (Mahfudhoh & Muslimin, 2020) revealed that service quality contributed only 26% to patient satisfaction, indicating that other factors play a significant role, underscoring the need for a deeper understanding of public perceptions.

RSUD Soe, as one of the Regional Government Organizations (OPD), has the vision of “Realizing High-Quality, Affordable, Fair, and Equitable Healthcare Services.” In efforts to realize this vision, RSUD Soe has implemented various initiatives to improve healthcare service quality. However, there are still complaints and grievances from the public, indicating that service performance has not yet reached optimal levels. These complaints are often conveyed through social media or directly. Therefore, it is important to understand public perceptions and sentiments toward the healthcare services provided by RSUD Soe.

To obtain a more comprehensive overview of public sentiment toward RSUD Soe services, the analysis was conducted using several machine learning methods, which have been widely applied in healthcare data analysis due to their effectiveness in handling large and complex datasets (Chen *et al*., 2017). In addition, classical machine learning algorithms are particularly suitable for text classification tasks with relatively limited datasets, as they are computationally efficient and have demonstrated robust performance in sentiment analysis (Manning *et al.*, 2008). The algorithms used include Naïve Bayes, which operates based on probability and the assumption that features are mutually independent (Alpaydin, 2020); C4.5, a decision tree algorithm that divides data based on the information gain ratio to construct classification structures (Salzberg, 1994). Random Forest is an ensemble learning method that builds multiple decision trees and combines their results through majority voting to enhance model accuracy and stability (Breiman L., 2001); and Support Vector Machine (SVM), an algorithm that seeks the optimal hyperplane to separate data classes with maximum margin (Cortes & Vapnik, 1995).

In addition to sentiment classification, this study also applies the Latent Dirichlet Allocation (LDA) method, an unsupervised topic modeling technique used to identify main themes, which is applied to uncover key topics within public suggestions or feedback. LDA views each document as a combination of various topics, with each topic consisting of a collection of words that have a high probability of appearing together (Blei *et al.*, 2003). This method has been widely used to extract hidden themes from healthcare-related textual data (Paul & Dredze, 2014). Through this approach, additional insights are expected to be obtained regarding the service aspects most frequently highlighted by the public.

This study aims to identify and analyze public perceptions and sentiments toward healthcare services at RSUD Soe, as well as to evaluate the performance of several machine learning methods in classifying such sentiments. In addition, this research seeks to explore the main topics within public suggestions using LDA. The results of the overall analysis are used as a basis for formulating relevant recommendations to improve the quality of healthcare services at RSUD Soe.

RESEARCH METHOD

This study is a descriptive study with a quantitative approach. Prior to data collection, the research population and sample were determined. The population in this study consisted of all patients who were currently undergoing or had previously received treatment at RSUD Soe during the period from September 2024 to February 2025. A purposive sampling technique was applied to select respondents based on specific criteria aligned with the research objectives, ensuring that only relevant participants were included in the study (Etikan et al., 2016). The criteria included patients who were currently receiving or had previously received treatment at RSUD Soe during the research period.

This study employed a closed-ended questionnaire based on a Likert scale with five response options, designed to measure patient perceptions and satisfaction levels regarding various aspects of healthcare services.

The questionnaire was developed based on indicators relevant to patient satisfaction, including SERVQUAL (Parasuraman *et al*., 1988), which encompasses tangibles, reliability, responsiveness, assurance, and empathy; satisfaction with waiting time (Andaleeb, 2001); perceptions of cleanliness and physical facilities (Donabedian, 1988); quality of communication and interaction with medical personnel (Stewart, 1995); service accessibility (Aday & Andersen, 1974); and perceptions of cost appropriateness relative to service quality (Zeithaml, 1988). These indicators were used to comprehensively capture patients' perceptions and overall satisfaction with healthcare services at RSUD Soe.

The research flow diagram can assist readers in better understanding the sequence of the research process, as illustrated in Figure 1.

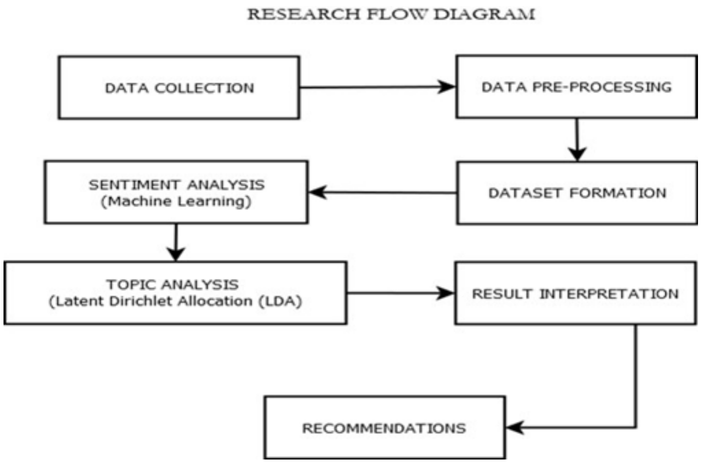

Figure 1. Research Flow Diagram.

Based on **Figure 1**, the research workflow begins with data collection by distributing questionnaires to members of the public who utilize the services of RSUD Soe. The questionnaire consists of several sections covering various areas of healthcare services and allows respondents to provide suggestions or feedback.

Data Pre-processing

Data pre-processing is the initial stage in data handling, aimed at cleaning and preparing the data before it is used in machine learning modeling. Pre-processing was applied to public response data collected through questionnaires. The pre-processing steps adopted in this study follow common methods in text mining and Natural Language Processing (NLP) (Witten *et al.*, 2016), with implementation supported by Indonesian-language NLP libraries such as Sastrawi. In this study, text data pre-processing was conducted through several essential stages. First, text cleaning was performed to remove text and numerical noise, punctuation marks, symbols, URLs, emoticons, and excessive whitespace in order to produce cleaner data. Next, case folding was applied to convert all letters to lowercase to ensure consistency. The tokenization stage segmented the text into individual words (tokens) to facilitate analysis. Finally, stemming was carried out to reduce words to their root forms using the Sastrawi stemmer, a library specifically designed for the Indonesian language (Nazief & Adriani, 2004).

After all stages were completed, the text became cleaner, more concise, and more focused on key terms relevant to respondents' sentiments. The results of each pre-processing stage are presented in Table 1.

Table 1. Results of data preprocessing

Original Text	Sentiment	Preprocessing Results
"Continue to work with sincerity and be friendly toward patients and visitors."	Positive	['continue', 'sincerity', 'friendly', 'patients', 'visitors']
"I suggest that services for patients be improved; room cleanliness and bathroom hygiene must be maintained, soap and towels should be available for handwashing, and the hospital environment should also be kept clean."	Negative	['suggest', 'service', 'patients', 'cleanliness', 'room', 'bathroom', 'clean', 'soap', 'towel', 'wash', 'hands', 'environment', 'hospital', 'clea
"Maintain and improve."	Neutral	[['maintain', 'improve']

The outcome of this pre-processing stage is cleaner text that is more focused on essential keywords, with a reduction in the number of unique words from approximately 2,500 to 850. This reduction enhances the efficiency and accuracy of the model in analyzing respondents' sentiments (Setiawan & Kusuma, 2020).

Dataset Labeling

To minimize subjectivity in the data annotation process, labeling was conducted by referring to systematically developed sentiment classification guidelines. These guidelines were designed to ensure consistency and validity in label assignment, as recommended by (Mohammad, 2016), which emphasize the importance of using clear guidelines in manual annotation processes.

The labeling was based on respondents' answers to the healthcare service satisfaction questionnaire at RSUD Soe. Sentiments were classified into three categories, namely Positive, Neutral, and Negative, based on the expressed level of satisfaction reflected in respondents' responses. The resulting labels were stored in the Kepuasan_Keseluruhan column and used as the target variable in sentiment classification analysis using machine learning algorithms.

Dataset Construction

The dataset used in this study was derived from patient questionnaire responses from individuals who received treatment at RSUD Soe. The dataset consists of 278 entries with a total of 16 columns representing various aspects of hospital services as well as patient suggestions. Dataset construction was carried out to prepare the data prior to conducting sentiment analysis and classification model training.

Input Columns and Labels

The input features used in the sentiment analysis process are a combination of several columns representing service aspects at RSUD Soe. These columns include Proses_Pendaftaran, Kejelasan_Informasi, Keramahan_Staf, Kemudahan_Pembayaran, Ketepatan_Waktu_Staf, Kualitas_Pelayanan_Medis, Waktu_Tunggu, Kebersihan_Kenyam anan, Fasilitas_Penunjang, Kejelasan_Medis, Keramahan_Medis, Kemudahan_Akses, and Kesesuaian_Biaya. All of these columns contain respondents' sentiment evaluations of each hospital service element. The values in each of these columns were then combined into a single narrative text that served as input for sentiment classification.

Meanwhile, the label or target for this classification was taken from the Kepuasan_Keseluruhan column, which reflects patients' overall assessment of the services provided. This label was divided into three categories—Positive, Neutral, and Negative—according to the sentiments expressed by the patients.

Example of Dataset Structure

After merging all service-related columns into a single text representation, each record produced a combined sentence referred to as Teks_Layanan. This column contains the textual representation derived from sentiment values of various service aspects, such as registration procedures, clarity of information, staff friendliness, and cost appropriateness. Meanwhile, the Kepuasan_Keseluruhan column serves as the final sentiment label used in the classification process. Figure 2 illustrates the dataset structure after the feature integration process.

	Teks_Layanan	Kepuasan_Keseluruhan
0	positif positif positif positif positif positi...	Positif
1	netral netral netral netral netral netral netr...	Netral
2	netral netral netral netral netral netral netr...	Netral
3	netral netral netral netral negatif netral neg...	Netral
4	netral netral netral netral negatif netral neg...	Negatif

Figure 2. Research Flow Diagram.

Dataset File Format

The dataset used in this study was stored in CSV (Comma-Separated Values) format with two main columns, namely Teks_Layanan and Kepuasan_Keseluruhan. The Teks_Layanan column contains the combined results of sentiment values derived from several service-related columns, while the Kepuasan_Keseluruhan column represents the final sentiment label reflecting the level of patient satisfaction. The CSV format was selected due to its compatibility with various software tools such as RapidMiner, Python (Pandas, Scikit-learn), and Weka. In this study, data processing and analysis were conducted using Google Colab.

Data Selection Criteria

In this study, only data entries with complete information in the service-related columns and the Kepuasan_Keseluruhan column were analyzed. Records containing missing values in these critical columns were removed to maintain the quality of the analysis. The Saran_Pasien column was not used in the primary sentiment classification and was separated for topic analysis using the Latent Dirichlet Allocation (LDA) method. This approach ensures that sentiment classification remains focused on service evaluations, while patient suggestions are analyzed separately to obtain additional insights.

Data Balancing

The initial distribution of sentiment labels in the Kepuasan_Keseluruhan column was imbalanced, with 142 Positive, 97 Neutral, and 39 Negative data points. To address this issue, oversampling techniques were applied to the Neutral and Negative classes until each class contained 142 data points. As a result, the class distribution became balanced, supporting more equitable and representative model training without requiring further balancing.

Sentiment Classification

The Kepuasan_Keseluruhan column was used as the label (target) for sentiment classification. The analysis was conducted using four machine learning algorithms: Naïve Bayes, C4.5, Random Forest, and Support Vector Machine (SVM). These algorithms are widely used in research and sentiment analysis and have been shown to be effective in text classification tasks (Pang & Lee, 2008; Wibowo & Saputra, 2021).

Algorithms Used

The selection of Naïve Bayes, C4.5, Random Forest, and Support Vector Machine (SVM) in this study is based on their respective strengths in handling text classification tasks, particularly for datasets with relatively limited size.

Naïve Bayes is a probabilistic classification method based on Bayes' Theorem, with the assumption that each feature is independent of the others (McCallum & Nigam, 1998).

$$P(C|X) = \frac{P(C)\,.P(X|C)}{P(X)}$$

C4.5 is a decision tree algorithm that uses the gain ratio to select the best attribute (Quinlan, 1993). The entropy and information gain are defined as :

$$Entropy(S) = -\sum p_i\, log_{2(p_i)}$$

$$Gain(S,A) = Entropy(S) - \sum \frac{|S_v|}{|S|} Entropy(S_v)$$

Random Forest is an ensemble method that constructs multiple decision trees and combines their prediction results using a voting technique (Breiman, 2001) :

$$\hat{y} = mode(\{\, h_{1(x)}, h_{2(x)}, \dots, h_{T(x)\}})$$

Support Vector Machine (SVM) is a classification algorithm that seeks the optimal hyperplane to separate data classes (Cortes & Vapnik, 1995).

$$\frac{1}{2}||w||^2 \; subject\ to\ y_{i(w\,\cdot x_i + b)} \geq 1$$

Although deep learning approaches such as LSTM and BERT have shown superior performance in large-scale text classification tasks, they typically require substantial datasets and higher computational resources. Given that this study utilizes a relatively limited dataset (278 respondents), classical machine learning algorithms are considered more appropriate and efficient for this context.

In this study, the parameter settings for each algorithm were defined as follows. The Naïve Bayes classifier was implemented using default parameters. The Decision Tree model used the entropy criterion with a fixed random state (random_state = 42). The Random Forest model was implemented using 100 trees with a fixed random state (random_state = 42). The Support Vector Machine (SVM) employed a linear kernel. These parameter settings were selected based on commonly used configurations to ensure a balance between model performance and computational efficiency.

Model Evaluation and Accuracy Testing

The balanced dataset was divided using a stratified train–test split approach with a proportion of 80% training data and 20% testing data to ensure that class distributions were preserved in both subsets (Kohavi, 1995). This proportion was selected as it represents a commonly used approach in machine learning, providing sufficient data for model training while maintaining a representative portion for reliable performance evaluation, particularly for datasets with limited size.

Each model was trained using the training dataset and evaluated on the testing dataset using several performance metrics derived from the confusion matrix. These metrics include accuracy, precision, recall, and F1-score, which are commonly used to assess classification performance (Powers, 2011).

Accuracy is defined as the proportion of correctly classified instances over the total number of instances:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision and recall are used to evaluate the model's performance for each class:

$$Precision = \frac{TP}{TP + FP} \; ; Recall = \frac{TP}{TP + FN} \; ;$$

The F1-score is the harmonic mean of precision and recall:

$$F1 - Score = 2 \cdot \frac{Precision \; \cdot Recall}{Precision + Recall}$$

These evaluation metrics provide a comprehensive assessment of model performance, particularly in multi-class classification tasks, by considering both the correctness of predictions and the balance between precision and recall across different sentiment classes.

Topic Analysis and Interpretation

O further explore public opinions and feedback, topic analysis was conducted using Latent Dirichlet Allocation (LDA). LDA is an unsupervised learning technique that groups documents into multiple topics based on word distribution patterns (Blei *et al.*, 2003). In this study, topic analysis focused on the open-ended input column containing respondents' comments and suggestions. The LDA results reveal the main topics that frequently emerge in public opinions. Interpreting these results helps identify dominant factors influencing satisfaction and supports the formulation of data-driven recommendations for RSUD Soe (Syahputra & Ruldeviyani, 2020).

The LDA model defines a generative process in which, for each document d, a topic distribution $\theta_{(d)}$ ~ Dirichlet (α) is first drawn. Then, for each word in the document, a topic $z_{(dn)}$ ~ Multinomial ($\theta_{(d)}$) is selected, followed by the selection of a word $w_{(dn)}$ ~ Multinomial ($\phi_{(z(dn))}$), where ϕ represents the word distribution for each topic (Blei *et al.*, 2003). The likelihood function of the corpus can be expressed as:

$$P(W \mid \alpha, \beta) = \prod_{d=1}^{D} \int P(\theta_d \mid \alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} P(z_{dn} \mid \theta_d) P(w_{dn} \mid z_{dn}, \beta) \right) d\theta_d$$

RESULTS AND DISCUSSION

This section presents the research findings and their discussion in a structured manner. The data collected were used to address the research questions and objectives. A total of 278 respondents participated in this study by evaluating the health services at Soe Regional General Hospital. The discussion in this section will focus directly on explaining the analytical model used and interpreting the results.

Sentiment Analysis Using Machine Learning

To determine the best model for classifying public sentiment toward the services of RSUD Soe, experiments were conducted using four machine learning algorithms, namely Naïve Bayes, Decision Tree (C4.5), Random Forest, and Support Vector Machine (SVM), with a stratified 80:20 split between training and testing data. The evaluation was performed based on accuracy, precision, recall, and F1-score metrics.

The parameter settings for each algorithm were determined based on commonly used default configurations and supported by previous studies in text classification, with minor adjustments made by the researchers to ensure a balance between model performance and computational efficiency.

1. Naïve Bayes

The Naïve Bayes model achieved an accuracy of 82.14%. Table 2 presents the complete performance of the Naïve Bayes model.

Table 2. Model *Naïve Bayes*

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.71	1.00	0.83	5
Neutral	0.79	0.85	0.81	26
Positive	0.90	0.76	0.83	25

The results presented in Table 2 indicate that the Naïve Bayes model demonstrates very strong performance in sentiment classification, particularly for negative sentiment, with a recall value of 1.00, indicating that all negative comments were correctly classified. Naïve Bayes is a probabilistic classification algorithm that assumes independence among features. This assumption is highly suitable for short text data derived from the RSUD Soe questionnaire, as each word tends to carry an independent meaning. The model calculates word probabilities for each sentiment class, enabling it to effectively recognize patterns even when negative data are fewer in number compared to other classes. Its efficiency, simplicity of implementation, and ability to handle imbalanced data make Naïve Bayes superior in this experiment.

2. Decision Tree (C4.5)

The Decision Tree (C4.5) model achieved an accuracy of 73.21%. Table 3 presents the complete performance of the Decision Tree (C4.5) model.

Based on the analysis of the performance tables that have been conducted, it is evident that the Decision Tree (C4.5) model provides a less optimal level of accuracy compared to the other models. Specifically, for negative sentiment, the F1-score achieved by this model is only 0.55 (Table 3). C4.5 constructs a decision tree structure by relying on information gain values. However, when the data contains a large number of word features with high similarity across classes or features that provide limited discriminatory information, the model is prone to underfitting. This condition may explain why C4.5 performs poorly in identifying negative comments within the RSUD Soe dataset, which is most likely due to the limited number of negative samples and the distribution of words that is insufficient to produce clear distinctions.

Table 3. Model Decision Trees

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.50	0.60	0.55	5
Netral	0.78	0.69	0.73	26
Positive	0.74	0.80	0.77	25

3. Random Forest

The Random Forest model achieved an accuracy of 79%. Table 4 presents the complete performance of the Random Forest model.

Table 4. Model Random Forest

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.50	1.00	0.67	5
Neutral	0.85	0.65	0.74	26
Positive	0.85	0.88	0.86	25

Based on the results presented in the previous table, the Random Forest model constructs multiple decision trees using subsets of data and features, and then averages their predictions. This approach enhances model stability and reduces the likelihood of overfitting. In the RSUD Soe dataset, Random Forest achieved perfect recall for the negative class (1.00), but its precision was low (0.50), indicating that many predicted negative instances were actually incorrect (Table 4). This may have occurred because the model was overly aggressive in labeling comments as negative, which can lead to a loss of overall accuracy.

4. Support Vector Machine (SVM)

The Support Vector Machine (SVM) model achieved an accuracy of 80%, with balanced performance across all classes. Table 5 presents the complete performance of the Support Vector Machine (SVM) model.

Table 5. Model Support Vector Machine (SVM)

Sentiment	Precision	Recall	F1-Score	Support
Negative	0.62	1.00	0.77	5
Neutral	0.83	0.77	0.80	26
Positive	0.83	0.80	0.82	25

The SVM performance table demonstrates strong results across all categories, with a recall of 1.00 for the negative class and a high F1-score for the positive class. SVM operates by projecting data into a higher-dimensional space and then identifying the optimal hyperplane to separate classes. Its primary advantage lies in its ability to recognize complex patterns, which is particularly beneficial for analyzing public opinion data such as that from RSUD Soe. However, SVM requires greater computational resources compared to models such as Naïve Bayes; therefore, efficiency considerations become important as data size increases.

5. Comparison of Model Accuracy

The Support Vector Machine (SVM) model achieved an accuracy of 80%, with balanced performance across all classes. Table 6 presents the complete performance of the Support Vector Machine (SVM) model.

Table 6. Comparison of Model Accuracy

Model	Accuracy
Naïve Bayes	82.14%
Decision Tree (C4.5)	73.21%
Random Forest	79%
Support Vector Machine (SVM)	80%

Based on the results of the analysis of the four machine learning algorithms applied in classifying public perceptions of healthcare services at RSUD Soe, it can be observed that the Naïve Bayes model achieved the best performance, with the highest accuracy of 82.14%. In contrast, other models, such as Support Vector Machine, attained an accuracy of 80%, Random Forest reached 79%, and Decision Tree (C4.5) obtained the lowest accuracy at 73.21% (Table 6).

In addition to its high accuracy, the Naïve Bayes model demonstrates advantages by effectively handling three sentiment categories (negative, neutral, and positive) and exhibiting a more balanced relationship among precision, recall, and F1-score values. In particular, for the negative sentiment category, the recall value reached 1.00, indicating that all data with negative sentiment were correctly classified. This aspect is especially important, as negative sentiment data are typically fewer in number and often pose challenges in the classification process.

The Naïve Bayes model is considered an effective method for text data processing because it relies on basic probability concepts and assumes feature independence. Although this assumption is not always accurate in real-world contexts, the model is capable of producing efficient and accurate classifications, particularly for short text data such as public reviews or questionnaire responses (Manning *et al*., 2008; Zhang, Lei *et al*., 2010).

Meanwhile, although the Random Forest and SVM models produced fairly good results, there were imbalances between precision and recall values in certain classes. In the Random Forest model, the precision value for the negative sentiment category was only 0.50 despite achieving perfect recall, indicating misclassification of data from other classes as negative. The Decision Tree (C4.5) model exhibited the weakest performance, particularly for negative sentiment, with an F1-score of only 0.55, reflecting its limited ability to detect negative comments.

Considering accuracy, performance stability, ease of implementation, and model efficiency, Naïve Bayes is the most appropriate method for classifying public sentiment toward healthcare services at RSUD Soe.

The Naïve Bayes model provides high accuracy in classifying public opinions regarding healthcare services at RSUD Soe. However, it also has several important limitations. One of the main issues lies in the assumption that all features are independent, where each word in a document is assumed to have no relationship with the sentiment class.

Topic Analysis Using Latent Dirichlet Allocation (LDA)

Topic analysis was conducted to identify the main themes emerging from public feedback and suggestions regarding healthcare services at RSUD Soe. The method employed was Latent Dirichlet Allocation (LDA), using an unsupervised learning approach to explore hidden patterns within the text data.

1. Selection of The Number of Topics Based on Cohenrence Score

To determine the most appropriate number of topics in Latent Dirichlet Allocation (LDA) modeling, an analysis was performed by comparing coherence score values across various numbers of topics, ranging from 2 to 15. Coherence values were used to measure the consistency and semantic interpretability of keywords within a topic, where higher values indicate topics that are more coherent and easier to understand.

The selection of the number of topics based on coherence scores aligns with the view of (Röder *et al.*, 2015), which states that the coherence score is an effective metric for evaluating topic quality in LDA. This metric can be used to determine the optimal number of topics to generate more meaningful and non-overlapping topics. Furthermore, (Blei et al., 2003), as the original developer of LDA, also emphasizes the importance of selecting an appropriate number of topics to ensure that the model can maximally represent the data. The evaluation results are presented as follows:

Table 7. Comparison of Model Accuracy

Number of Topics	Coherence Score
2	0.3248
3	0.3172
4	0.3521
5	0.3567
6	0.3601
7	0.3638
8	0.3578
9	0.4042
10	0.3875
11	0.3408
12	**0.4160**
13	0.3643
14	0.3563
15	0.3931

Based on Table 7, the highest coherence score was obtained when the number of topics was set to 12, with a value of 0.4160. Therefore, topic modeling in this study was conducted using 12 topics, which were considered the most optimal.

To strengthen the results of topic number selection, an Intertopic Distance Map visualization was generated using LDAvis for the model with 12 topics. As shown in Figure 3, each circle in the visualization represents a single topic, and the size of the circle indicates the proportion of that topic within the overall review dataset. The distance between circles reflects the semantic differences among topics. Many topics are positioned far apart, indicating that each topic captures distinct issues with minimal overlap. This further reinforces the conclusion that the selection of 12 topics provides an appropriate and representative modeling of the data.

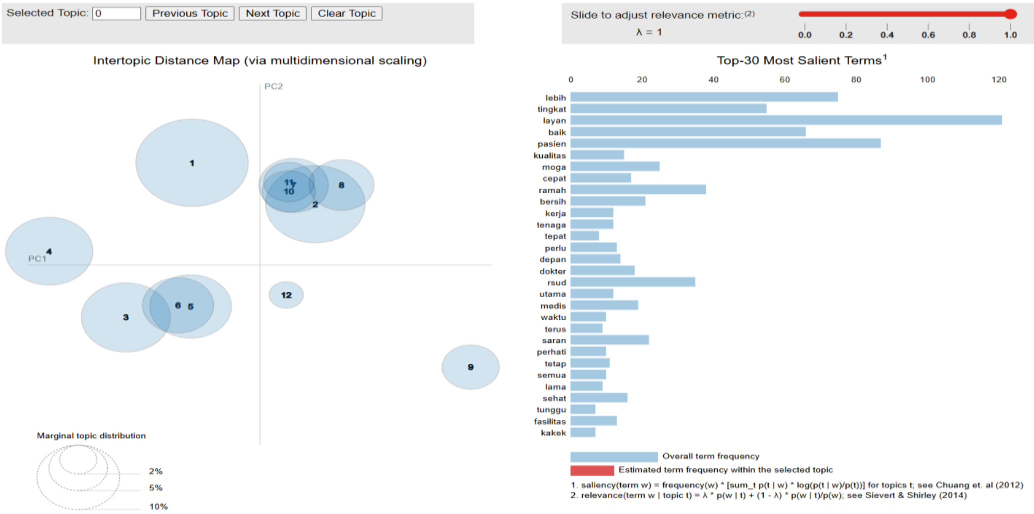

Figure 3. LDA Topic Visualization Using pyLDAvis

The right side of Figure 3 displays a list of the 30 most salient terms (Top-30 Most Salient Terms) based on their frequency and relevance across all topics. Words such as layan (service), baik (good), pasien (patient), ramah (friendly), and cepat (fast) appear with high frequency, indicating that most reviews focus on service quality, staff attitudes, and service speed. These terms demonstrate that the generated topics are meaningful and aligned with the research context of hospital service sentiment analysis.

In addition, LDAvis provides a useful parameter, lambda (λ), which controls the balance between term frequency and term specificity within each topic. When λ is high, the visualization highlights the most frequent words, while lower λ values reveal more unique or distinctive terms for each topic. This helps ensure a more comprehensive interpretation of topic characteristics.

Thus, supported by coherence scores, LDAvis visualization, and the interpretability adjustment through lambda, it can be concluded that the model with 12 topics provides the best representation of the topic structure in patient review data at RSUD Soe.

Topic Results and Interpretation

The following is a list of the 12 topics generated by the LDA model, along with the dominant keywords for each topic.

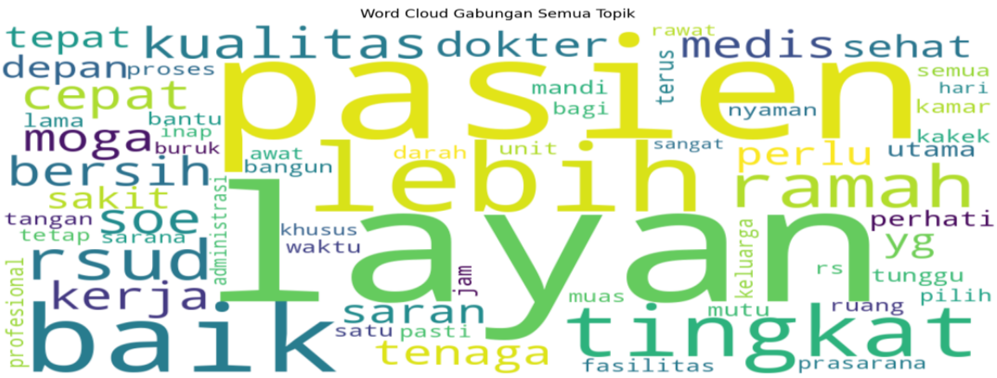

Figure 4. Combined Word Cloud of All Topics

The LDA model generated 12 main topics related to services at RSUD Soe based on public opinions. One theme focuses on complaints regarding waiting times and expectations for faster services delivered with a friendly attitude. There is also an expectation that the overall quality of services at RSUD Soe will continue to improve.

Attention to specific patient groups, particularly the elderly, as well as the need for clean and comfortable facilities, also received considerable emphasis. The community highlighted the importance of having an adequate number of responsive medical personnel in patient care. In addition, comfort during the service process is expected to receive greater attention, with hopes for more personalized and humane services. The public expressed a desire for RSUD Soe to improve its services in the future. Service quality and accuracy are considered highly important, indicating a need for professionalism and efficiency among healthcare workers. There were also suggestions to improve physical facilities, infrastructure, blood availability, and operational schedules (Figure 4).

The attitude of healthcare workers, characterized by professionalism and kindness, is considered crucial. The community expects services to be fast and responsive. Service speed and cleanliness are highly emphasized, indicating that efficiency must continue to prioritize patient comfort. Furthermore, services in the Emergency Department (IGD) and inpatient wards require particular improvement, as these areas are a primary focus of public concern.

Recommendations Based on Topic Analysis Results

Based on the results of topic exploration using the LDA method, several key insights were identified as a foundation for improving healthcare services at RSUD Soe. The findings reveal that service quality and speed remain major concerns, as many respondents highlighted long waiting times and expressed expectations for faster, friendlier, and more consistent services. Therefore, it is necessary to evaluate service workflows, enhance healthcare worker training in service excellence and patient communication, and implement more transparent and efficient queuing systems. In addition, the need for the addition and equitable distribution of medical personnel was emphasized, suggesting that recruitment or reallocation of staff should be aligned with the specific needs of each service unit, supported by regular competency development, particularly for emergency and inpatient services.

Furthermore, issues related to supporting facilities, including waiting rooms, inpatient wards, service units, and blood availability, indicate the importance of routine maintenance and infrastructure improvement to enhance patient comfort. This can be achieved by reorganizing spaces according to comfort standards and providing adequate supporting facilities such as clean toilets, spacious waiting areas, and proper ventilation. Finally, patient satisfaction and comfort were consistently highlighted, underscoring the need for training in professional attitudes and empathy, the implementation of periodic patient feedback monitoring systems, and the maintenance of a clean, comfortable, and patient-friendly hospital environment.

Interpretation of Sentiment and Topic Analysis Results

This section presents an interpretation aimed at analyzing more deeply the relationship between public sentiments regarding services at RSUD Soe (positive, neutral, and negative) and the main issues identified through comment analysis using machine learning methods and topic modeling. This combined method is referred to as sentiment-aware topic modeling, which integrates emotional and thematic analysis simultaneously to achieve a more comprehensive and practical understanding for formulating service quality improvement strategies.

The results of sentiment analysis using four machine learning algorithms indicate that the Naïve Bayes algorithm demonstrated the best performance in identifying patient sentiments, achieving the highest accuracy compared to C4.5, Random Forest, and SVM algorithms. Based on the classification results, most respondents provided feedback with positive sentiment, particularly regarding staff friendliness, service quality, and response speed. Nevertheless, a substantial portion of comments reflected neutral and negative sentiments. Neutral comments were generally informative or provided suggestions without explicit emotional expression, while negative comments were mostly related to long waiting times, insufficient numbers of medical personnel, and limited physical facilities.

Topic analysis using the Latent Dirichlet Allocation (LDA) approach explains the context of these sentiments. Topics containing keywords such as good, friendly, and improved services were commonly found in comments reflecting positive sentiment, indicating public appreciation of the quality of interactions between patients and healthcare workers, as well as efforts to enhance services. Conversely, topics containing keywords such as long, waiting, lack of staff, and facilities appeared in negative comments, highlighting structural and operational issues. Meanwhile, topics with keywords such as patient, suggestion, and comfort were more frequently found in neutral comments, indicating opportunities for long-term improvements that merit further investigation.

This comprehensive interpretation indicates that public perceptions of RSUD Soe have multiple dimensions, encompassing both emotional elements and specific issues. By systematically identifying the relationship between sentiments and issues, RSUD Soe can prioritize service improvements more effectively. Issues that consistently appear in negative reviews, such as insufficient medical staff and limited facilities, can be designated as top priorities for improvement. On the other hand, positive aspects such as the friendly attitudes of medical personnel should be maintained and further enhanced.

This combined approach provides a holistic understanding of how the public feels and what they discuss. Sentiment analysis categorizes comments into positive, neutral, and negative based on emotions, while LDA topic analysis identifies key issues regardless of sentiment. Mapping sentiments to topics reveals the relationship between emotions and the issues discussed. For example, long waiting times and a lack of medical staff frequently appear in negative comments, while friendliness and good service are more prevalent in positive comments. This relationship facilitates a deeper understanding of public perceptions and can be used as a basis for decision-making to improve service quality.

Recommendations for RSUD Soe

Based on the results of sentiment classification and topic analysis of public comments regarding healthcare services at RSUD Soe, several strategic recommendations can be proposed to improve service quality. The findings indicate that long waiting times and service queues remain key issues; therefore, improvements in service speed and efficiency are necessary through the evaluation of patient registration and initial service systems, increasing staff during peak hours, and optimizing electronic queuing or automated calling systems. In addition, the attitudes and ethics of medical personnel play a crucial role in shaping positive sentiment, highlighting the need for regular training in communication and excellent service, as well as the provision of rewards or incentives for staff receiving positive patient feedback. From the facility perspective, maintaining cleanliness and improving hospital infrastructure, including inpatient rooms and public facilities, are essential, along with periodic evaluations of patient satisfaction. Furthermore, the adoption of technology, particularly sentiment analysis–based monitoring systems supported by visual dashboards, can assist in continuously evaluating public opinions and service performance. Special attention should also be given to vulnerable patient groups by providing priority service pathways and strengthening coordination across service units to ensure more responsive care.

The combined use of sentiment analysis and topic modeling demonstrates that data-driven approaches can provide deeper insights into public perceptions and needs. While sentiment analysis classifies responses into positive, neutral, and negative categories, topic modeling identifies the key issues underlying these responses. These approaches are complementary: sentiment classification reflects overall satisfaction levels, whereas LDA explains the reasons behind the sentiments, particularly in negative feedback that highlights issues such as waiting times, limited facilities, and medical staff attitudes. By identifying dominant topics within each sentiment category, especially negative ones, RSUD Soe can formulate more effective, targeted, and sustainable strategies to improve the overall quality of healthcare services.

CONCLUSION

Based on the research conducted on sentiment analysis of healthcare services at RSUD Soe, this study successfully identified and evaluated public perceptions of healthcare services at RSUD Soe by utilizing machine learning approaches and Latent Dirichlet Allocation (LDA). The data used were obtained from 278 respondents and underwent text preprocessing and classification using four algorithms: Naïve Bayes, Decision Tree (C4.5), Random Forest, and Support Vector Machine (SVM). Based on model performance evaluation results, the Naïve Bayes algorithm demonstrated the best performance in classifying public sentiment, achieving the highest accuracy of 82.14%, followed by SVM (80%), Random Forest (79%), and C4.5 (73.21%). The Naïve Bayes model also excelled in detecting negative sentiment with a perfect recall value (1.00), making it the most appropriate method for this analysis. Topic analysis using LDA generated 12 main topics reflecting the issues most frequently highlighted by the public, such as long waiting times, shortages of medical personnel, cleanliness and physical facilities, and healthcare worker attitudes. Visualization and coherence scores supported that the 12-topic model optimally represented the data. By combining sentiment and topic analysis results, a comprehensive understanding of public perceptions was obtained. Most comments were positive, but there were also important neutral and negative inputs that can serve as a basis for service improvements. Overall, this study demonstrates that data-driven and machine learning-based modeling approaches can be effectively used to evaluate healthcare service quality and formulate more measurable and sustainable strategic recommendations for improving service quality at RSUD Soe.

ACKNOWLEDGMENT

The authors respectfully express their gratitude to their supervisors, the Soe Institute of Education, Satya Wacana Christian University, and RSUD Soe for all guidance, direction, support, and facilities provided during the completion of this research.

REFERENCES

Abd-alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study Corresponding Author. *Journal of Medical Internet Research, 22(4)*, e19016. https://doi.org/10.2196/19016

Aday, L. A., & Andersen, R. (1974). A framework for the study of access to medical care. *Health Services Research*, *9*(3), 208–220.

Alpaydin, E. (2020). Introduction to Machine Learning (4th ed.). The MIT Press.

Andaleeb, S. S. (2001). Service quality perceptions and patient satisfaction: A study of hospitals in a developing country. *Social Science \& Medicine*, *52*(9), 1359–1370. DOI: 10.1016/s0277-9536(00)00235-5

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022. DOI: 10.5555/944919.944937

Breiman, L. (2001). Random forest. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease Prediction by Machine Learning Over Big Data From Healthcare Communities. *IEEE Access*, *5*, 8869–8879. https://doi.org/10.1109/ACCESS.2017.2694446

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Donabedian, A. (1988). The quality of care: {How} can it be assessed? *JAMA*, *260*(12), 1743–1748. DOI: 10.1001/jama.260.12.1743

Etikan, I., Musa, S.A., & Alkassim, R.S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., & Donaldson, L. (2013). Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. *BMJ Quality & Safety*, *22*(3), 251–255. https://doi.org/10.1136/bmjqs-2012-001527

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, 1137–1145. https://doi.org/10.5555/1643031.1643047

Lai, S. T., & Mafas, R. (2022). Sentiment Analysis in Healthcare: Motives, Challenges & Opportunities pertaining to Machine Learning, *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, 8(1), 1–4. https://doi.org/10.1109/ICDCECE53908.2022.9792766

Mahfudhoh, M., & Muslimin, I. (2020). Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pasien Pada Rumah Sakit Umum Daerah Kota Cilegon. *JIMKES*, *8*(1), 39–46. https://doi.org/10.37641/jimkes.v8i1.310

Manning, C. D., Raghavan, P., & Schütze, H. (2008). An Introduction to Information Retrieval. Cambridge University Press.

McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. *AAAI Conference on Learning Artificial Intelligence*.

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. *Emotion Measurement,* 201–237). https://doi.org/10.48550/arXiv.2005.11882

Mohd Sofi, S., & Selamat, A. (2023). Aspect Based Sentiment Analysis: Feature Extraction using Latent Dirichlet Allocation (LDA) and Term Frequency - Inverse Document Frequency (TF-IDF) in Machine Learning (ML). *MyJICT*, 169–179. https://doi.org/10.53840/myjict8-2-102

Nazief, B., & Adriani, M. (2004). Stemming algorithm for Bahasa Indonesia. *International Conference on Information Technology and Multimedia*.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, *2*, 1–135. https://doi.org/10.1561/1500000011

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.

Paul, M.J., & Dredze, M. (2014). Discovering Health Topics in Social Media Using Topic Models. *PLOS ONE*, *9*(8), 1–11. https://doi.org/10.1371/journal.pone.0103408

Powers, D.M.W. (2011). Evaluation: From precision, recall and F-measure to ROC, *informedness, markedness & correlation*. *2*(1), 37–63. https://doi.org/10.48550/arXiv.2010.16061

Quinlan, J.R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.

Rao, J., Zhang, Q., Liu, S., & Liu, X. (2024). Machine Learning in Action: Topic-Centric Sentiment Analysis and Its Applications. *IJMRGE*, *5*(6), 1274–1278. https://doi.org/10.54660/.IJMRGE.2024.5.6.1274-1278

Republic of Indonesia. (2009). Law No. 36 of 2009 on Health.

Röder, M., Both, A., & Hinnerburg, A. (2015). Exploring the space of topic coherence measures. https://doi.org/10.1145/2684822.2685324

Salzberg, S. L. (1994). C4.5: Programs for Machine Learning by J. Ross Quinlan. *Machine Learning*, *16*(3), 235–240. https://doi.org/10.1007/BF00993309

Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., Upadhaya, T., & Gonzalez, G. (2015). Utilizing social media data for pharmacovigilance: A review. *Journal of Biomedical Informatics*, *54*, 202–212. https://doi.org/https://doi.org/10.1016/j.jbi.2015.02.004

Setiawan, H., & Kusuma, A. (2020). Penerapan Text Preprocessing untuk Analisis Sentimen pada Ulasan Pengguna. *Jurnal Teknologi Informasi Dan Komputer*, *6*(2), 45–53.

Sharma, A., & Dey, S. (2012). A comparative study of feature selection and machine learning techniques for sentiment analysis. *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, 1–7. https://doi.org/10.1145/2401603.2401605

Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE Journal of Biomedical and Health Informatics*, *22*(5), 1589–1604. https://doi.org/10.1109/JBHI.2017.2767063

Stewart, M. A. (1995). Effective physician-patient communication and health outcomes: A review. *CMAJ*, *152*(9), 1423–1433. PMCID: PMC1337906

Syahputra, R., & Ruldeviyani, Y. (2020). Penerapan LDA untuk Analisis Topik pada Ulasan Pengguna. *Jurnal Teknik ITS*, *9*(1), A86-A91.

Wibowo, A., & Saputra, R.E. (2021). Klasifikasi Sentimen dengan Metode Naive Bayes, Decision Tree dan SVM. *Jurnal RESTI*, *5*(1), 10–17.

Witten, I. H., Frank, E., & Hall, M.A. (2016). Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Elsevier Inc.

Zeithaml, V. A. (1988). Consumer perceptions of price, quality and value: A means–end model and synthesis of evidence. *Journal of Marketing*, *52*(3), 2–22. https://doi.org/10.2307/1251446

Zhang, Lei., Wang, Shuai., & Liu, Bing. (2010). Understanding sentiment analysis: Mining opinions, sentiments, and emotions. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *5*(1), 1–40.