Multilabel Aspect-Based Sentiment Analysis (M-ABSA) of Healthcare Services Based on Opinion Extraction and Powerset Labeling

Fika Hastarita Rachman
Informatics Department, University of Trunojoyo, Madura, Bangkalan, Indonesia
fika.rachman@trunojoyo.ac.id (corresponding author)

Firdaus Solihin
Informatics Department, University of Trunojoyo, Madura, Bangkalan, Indonesia
fsolihin@trunojoyo.ac.id

Ika Oktavia Suzanti
Informatics Department, University of Trunojoyo, Madura, Bangkalan, Indonesia
iosuzanti@trunojoyo.ac.id

Imamah
Department of Information Systems, University of Trunojoyo, Madura, Bangkalan, Indonesia
i2m@trunojoyo.ac.id

Deshinta Arrova Dewi
Faculty of Data Science and Information Technology, INTI International University, Malaysia
deshinta.ad@newinti.edu.my

Received: 15 December 2025 | Revised: 3 February 2026 and 26 February 2026 | Accepted: 15 March 2026

 DOI: https://doi.org/10.48084/etasr.16961

ABSTRACT

Healthcare facility services are important in supporting public health and reducing health risks. Multilabel Aspect-based Sentiment Analysis (M-ABSA) can be used to analyze patient reviews from social media, allowing healthcare facility managers to identify areas that need improvement based on public perception, thereby supporting the continuous enhancement of healthcare service quality. This study compares two M-ABSA approaches: a transformation approach with powerset labeling and feature engineering techniques with opinion extraction. In addition, a traditional machine learning method (Random Forest) is compared with deep learning methods such as Long Short Term Memory (LSTM) and Indonesian Bidirectional Encoder Representative from Transformers (IndoBERT). The experimental results show that powerset labeling with random oversampling led to an accuracy of 0.79, while M-ABSA using opinion extraction achieved 0.81 with LSTM. These results indicate that the feature engineering approach using opinion extraction combined with deep learning, particularly LSTM, performs well in analyzing healthcare facility user review data.

Keywords-natural language processing; multilabel aspect-based sentiment analysis; opinion extraction; powerset labeling; public health; health risk; product innovation; social protection system; deep learning

I. INTRODUCTION

Healthcare services directly affect the health and social integration of people in urban and rural areas [1, 2]. Improving the quality of healthcare services is one of the responsibilities of healthcare providers to ensure that community needs are met [3-6]. In this regard, efforts are being made to review the quality of healthcare services and patient safety [7]. The quality of health services and hospital performance has improved in various clinical and administrative aspects as a result of accreditation [8, 9], which requires the support of both clinical and non-clinical personnel [7].

The International Accreditation Standards of the Joint Commission International (JCI) [10], the National Hospital Accreditation Standards (SNARS) [11], and the World Health

Organization (WHO) [12] are references in the accreditation of healthcare facilities. There are several assessment domains, including patient-centered care, governance, environment & safety, human resources, information management, quality & safety, clinical support, and finance. According to standards, there are many factors in accreditation assessments. However, not all of these factors are reviewed by patients on social media. Since each visitor has a different visit time, and some only receive outpatient or inpatient care, not all patients may be aware of all aspects of hospital management and governance. Reviews on social media can be self-reported patient experiences with healthcare services [13, 14]. Self-reported data is very useful for healthcare facility managers to improve service quality, provided that they can extract and analyze reviews appropriately. It is crucial to extract the aspects discussed within numerous reviews on social media. However, the abundance of reviews on social media about healthcare facilities makes it difficult for service providers to determine which aspects need improvement and what needs to be addressed. To address this issue, text mining methods are needed to analyze and summarize public reviews on social media.

Sentiment analysis can be used to measure patient satisfaction with medical services and the quality of healthcare services using data from social media. Sentiment analysis has evolved to Aspect-Based Sentiment Analysis (ABSA) to extract the aspects discussed in social media reviews [15, 16]. This approach aims to address the problem of analyzing documents that consist of several different aspects. For example, a patient reviews the Hospital health facility service as "The doctor is good, but the facilities are not very supportive, so the treatment is not very good," which shows that the patient gave a positive review of the healthcare staff's services but a negative review of the hospital's facilities. A single review can contain multiple aspects with differing sentiments. This can be addressed using Multilabel Aspect-Based Sentiment Analysis (M-ABSA) [17].

The determination of aspects in M-ABSA research is carried out manually by each researcher. In [18], three aspects were identified, whereas in [19], 10 types were identified. Based on hospital accreditation standards and the seven dimensions of hospital service quality, the evaluation factors that can be directly assessed by the public are medical and staff service quality factors (effectiveness, safety, patient-centeredness quality dimensions), facilities and infrastructure (safety, efficiency dimensions), service waiting time (timeliness), and service costs (efficiency cost) [20]. In this study, reviews were mapped into four aspects, namely, quality of medical services and staff [20], facility and infrastructure [20, 21], service waiting time [20], and service cost [20, 21].

M-ABSA techniques are very diverse. Some studies applied a labeling transformation approach technique, Powerset Labeling (PL), which converts combinations of aspects and sentiments into unique labels [22, 23]. In [24], a feature engineering approach was applied to extract data prior to classification with Machine Learning (ML). This study aimed to compare different M-ABSA solution techniques.

The feature engineering technique used in this study is opinion extraction [25], which is a process to generate new information from primary reviews consisting of several sentences and with many aspects [24]. A review consists of several sentences, each discussing different aspects. The novelty of this study is the formation of opinion segments in the feature engineering process before the feature extraction stage. This segment formation is called "opinion extraction using a rule base." To extract opinion sentences and identify their patterns, one must use pattern knowledge and contextual information as references [26]. POS tagging is used as a tool to create the rule base for pattern knowledge in opinion extraction [27]. The review labeling process can be performed using expert-based and lexicon-based methods [28]. This study implements two approaches: expert validation for primary reviews and lexicon-based validation with TextBlob and VADER for opinions. The presence of a data imbalance also affects the accuracy of the system. ABSA requires a specific model to handle the cases of imbalanced datasets [29].

This study aimed to address the M-ABSA problem by applying opinion extraction, resolving data imbalance, and modeling using Deep Learning (DL) methods. The contributions of this research include:

- Establishes an appropriate and suitable rule base for opinion extraction to serve as the basis for the feature engineering process of converting original reviews into sub-reviews in the form of opinions.
- Integrates feature engineering (opinion extraction) techniques with traditional ML in a hierarchical manner within M-ABSA.
- Implements a PL approach, a method for handling imbalanced data in M-ABSA, with DL learning for healthcare facility reviews.
- Compares the labeling transformation approach with the feature engineering approach in M-ABSA, as well as the application of traditional ML with DL, to determine the best model for healthcare facility managers as evidence-based insights to improve service quality to support healthcare facility accreditation.

II. PROPOSED METHOD

This study developed two (2) systems: An M-ABSA using opinion extraction (M1) and an M-ABSA using PL (M2).

A. Dataset Formation

The data used in this study were scraped from Google Maps reviews for four healthcare facilities in East Java, Indonesia, resulting in 6,134 reviews in Indonesian. The review period was between December 2024 and August 2025. The next step involved data filtering based on duplicate data, reviews lacking sentiment or neutrality, and the language used. The filtering process resulted in 2,500 reviews of healthcare facilities. The average length of the reviews ranged from 34 to 35 words. After data collection, sentiment labeling was performed manually by analyzing the data and assigning aspect and sentiment labels, which were validated by Indonesian language experts. This study used four aspects, each with a positive (+) and negative (-) sentiment value:

- Aspect 1 (A1): Medical Service Quality & Staff
- Aspect 2 (A2): Facilities & Infrastructure
- Aspect 3 (A3): Waiting Time
- Aspect 4 (A4): Service Cost

Table I shows manual labeling for the following two review examples:

- Review 1 (R1): The hospital is clean, the polyclinic nurses are all friendly, especially the nurses are very friendly.
- Review 2 (R2): The service was excellent: friendly, informative, and patient. The hospital facilities were also excellent, complete, and clean. My only complaint is that the administration department spelled my name incorrectly.

TABLE I. EXAMPLE OF MANUAL LABELING FOR TWO REVIEW

Reviews	A1+	A1-	A2+	A2-	A3+	A3-	A4+	A4-
R1	1	0	1	0	0	0	0	0
R2	1	1	1	0	0	0	0	0

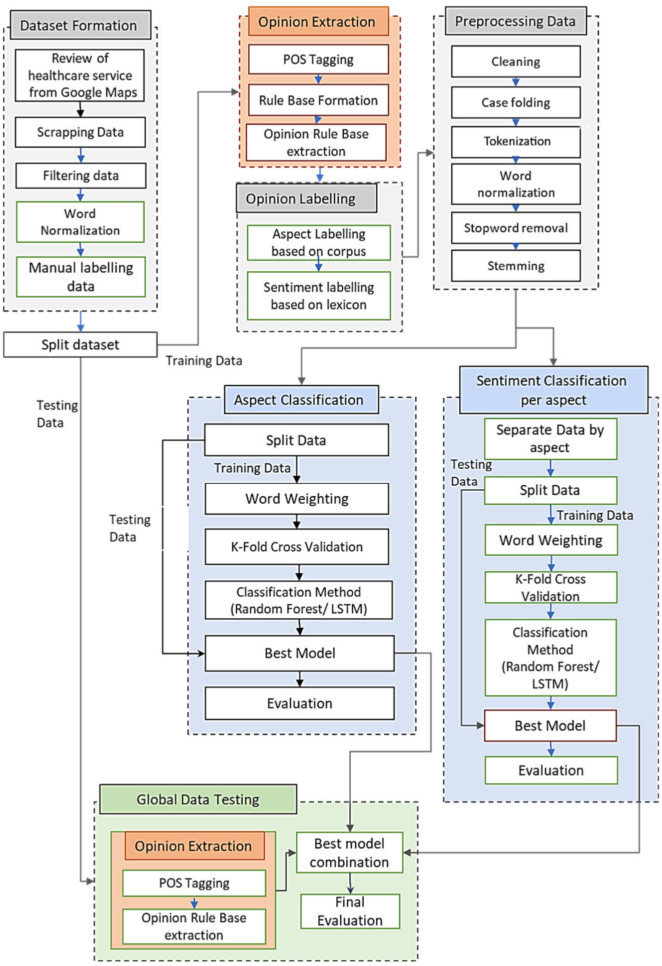

Fig. 1. Proposed system: M-ABSA using opinion extraction (M1).

B. M-ABSA Using Opinion Extraction (M1)

Figure 1 shows the M1 architecture. Health service reviews resulting from scraping Google Maps reviews and subsequent data filtering are input to the opinion extraction stage.

1) Opinion Extraction

In the opinion extraction process, a review is broken down into several opinion sentences, each of which may have different aspects and sentiments. The application of opinion extraction in M-ABSA can reduce the system's confusion in identifying opinion sentences as positive or negative sentiments on certain aspects.

An example of a review sentence is as follows: "The doctor was good, but the facilities were inadequate, so the patient was not handled well." The opinion extraction results from this review are:

- Opinion sentence 1: The doctor was good.
- Opinion sentence 2: The facilities were inadequate, so the patient was not handled well.

The two (2) opinion sentences resulting from the review extraction are more specific, indicating certain aspects and sentiments. This opinion extraction process is expected to improve the accuracy of the M-ABSA system. This process went through the following three stages.

- Part of Speech Tagging (POS-Tagging): The specific marking (tag) provided includes the word's suitability to the defined word class, indicating its grammatical category, such as noun, verb, or adjective, and so on. To carry out the tagging process, a tagset is required to group words. This study uses a dataset in Indonesian, so the tagset used is the NLP-idn tagset, which has 37 tags. The tagset resulting from the POS tagging will later be used as an opinion pattern that serves as a rule base to break down reviews into opinions.
- Rules' Formation: This process aims to find opinions contained in documents. Words, phrases, and written documents conceal opinions. Sentences that express a point of view are the smallest semantic units that can be used to obtain views. It is necessary to form rules using an Indonesian rule-based approach to determine the opinion patterns that will be generated. These rules are obtained by identifying the data used in the research.
- Rule-based Extraction: In this stage, the review is broken down into opinions based on the established rules. Rule formation is based on opinion sentence patterns that already have POS tagging and frequently appear in the training data. Figure 2 shows an example of rule formation from several POS-tagged sentences. Using the same concept as in Figure 2, 40 rules are formed, which will serve as a reference in the rule base extraction process. The results of opinion rule-based extraction on 2,500 training data items yielded 5,531 opinion sentences.

Reviews data:	1. Pelayanannya ramah, perawatnya ramah-ramah terutama di poli jasmine 😍 *(The service is friendly, the nurses are friendly, especially at the Jasmine polyclinic.)*
	2. Pelayanannya bagus dan baik Susternya sangat membantu dan baik-baik juga. *(The service is good and kind. The nurses are very helpful and kind too.)*
	3. Pelayanan baik banget, fasilitas lengkap besar, semuanya ramah, teliti, terbaik *(Very good service, large complete facilities, everyone is friendly, thorough, the best)*
Pos Tagging result:	1. **[Pelayanan/NN] [nya/DT] [ramah/JJ]** [,/SYM] [perawat/NN] [nya/PR] [ramah-ramah/JJ] [terutama/JJ] [di/IN] [poli/NN] [jasmine/FW] [😍/NNP]
	2. **[Pelayanan/NN] [nya/DT] [bagus/JJ]** [dan/CC] [baik/JJ] [Suster/NNP] [nya/PR] [sangat/ADV] [membantu/VB] [dan/CC] [baik-baik/JJ] [juga/ADV] [./SYM]
	3. **[Pelayanan/NN] [nya/DT] [baik/JJ]** [banget/ADV] [,/SYM] [fasilitas/NN] [lengkap/JJ] [besar/JJ] [,/SYM] [semuanya/ADV] [ramah/JJ] [,/SYM] [teliti/VB] [,/SYM] [terbaik/JJ]
Rule Based formation	**NN+DT+JJ**

Fig. 2. Example of Rules' formation.

A total of 40 rules was generated in this stage: 37 rules structured based on POS tagging, and 3 rules based on punctuation. Table II shows an example of 5 of the 40 rules used in opinion extraction. Table III shows an example of opinion extraction based on the rules formed.

TABLE II. EXAMPLE OF RULE-BASED FORMATION RESULTS

Structure of rules	Definition	Rule
Based on Stylistic	Rules created based on punctuation (symbols)	…+ **/SYM+**… SYM = Tag symbol (,.:)
Based on POS tagging	Rules created based on the POS Tagging pattern	If any (NN+JJ) or (NN+VB+JJ) or (NN+ADV+JJ) or (NN+PR+JJ) then opinion

NN=Noun; JJ=Adjective; VB=Verb; ADV=Adverb; PR=pronoun

TABLE III. EXAMPLE OF OPINION EXTRACTION

Reviews	Rule-based used	Opinion extraction results
R1	If any (NN+JJ), then opinion	The hospital is clean
		The polyclinic nurses are all friendly
	If any (NN+ADV+JJ), then opinion	especially the nurses are very friendly

2) *Opinion Labeling*

The opinion data are then automatically relabeled using a lexicon for aspect and sentiment labels. Therefore, each opinion will have an aspect label and a sentiment label. The aspect labels are determined using a keyword corpus for each aspect. The researchers created keywords based on words that frequently appear in reviews for a specific aspect. For sentiment labels, this study used Lexicon-based extraction, examining the use of TextBlob and VADER.

3) *Aspect Classification*

The opinion data was used in parallel as input for the classification of the aspects and the classification of sentiment per aspect. Classification was performed by dividing the data using K-fold cross-validation (K=5). This study compared traditional ML (Random Forest) with DL (LSTM) methods.

4) *Sentiment Classification per Aspect*

Sentiment classification for each aspect will likely differ, so this process was divided into four models: sentiment classification for aspects A1, A2, A3, and A4. To develop these models, opinion data were divided by aspect to serve as input for each classification. Classification was performed using Random Forest and LSTM.

5) *Global Data Testing*

In this stage, global testing was performed on the test data, which consisted of original reviews from Google Maps. The best models from aspect classification were combined with the ones from sentiment classification. Figure 3 shows the flowchart for the best model combination for final evaluation. The test results were evaluated to measure system performance.

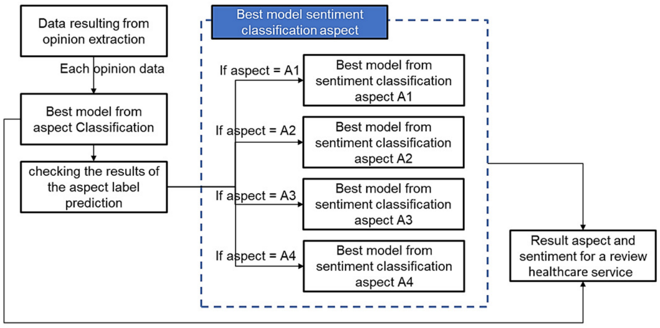

Fig. 3. The merging flow for the best model combination stage.

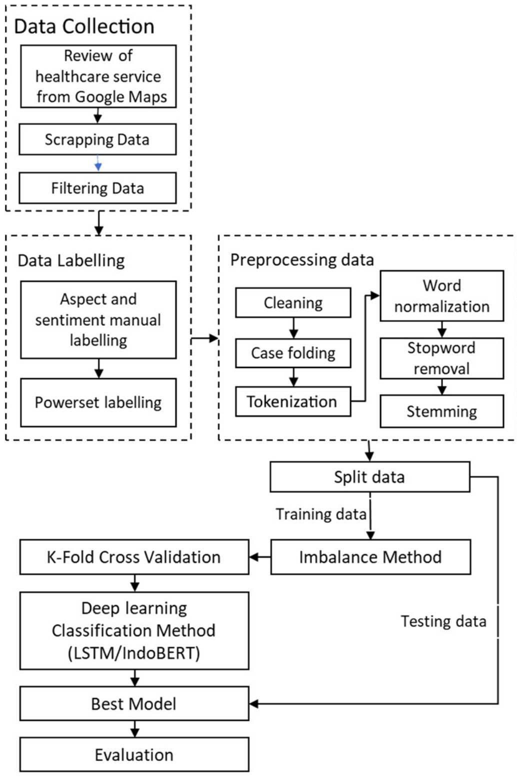

Fig. 4. Design system M-ABSA using Powerset Labeling (PL) (M2).

C. Design System M-ABSA Using Powerset Labelling (M2)

The research flowchart in Figure 4 illustrates the M-ABSA using PL (M2). The difference from M1 lies in the multilabeling technique; in M2, there is no opinion extraction process. Review data from Google Maps is included in the label transformation stage using PL.

1) Data Labelling

PL is a transformation labeling method for handling multi-label classification, which treats a combination of labels consisting of binary digits as a new class. PL transforms a multilabel case into a single label. Initially, the results of manual labeling for aspects and sentiments were transformed into PL. Table IV shows an example of the labeling transformation with PL. The number of unique data resulting from PL is 69 combination labels (C1 to C69). Of the entire PL set, the frequency distribution of the data is not the same. Thus, oversampling was used y for test scenarios with 10, 15, and 20 PLs. The selected powerset labels were those with the highest frequencies.

TABLE IV. EXAMPLE OF LABELING TRANSFORMATION WITH PL FOR 2 REVIEW DATA

Reviews data	A1+	A1-	A2+	A2-	A3+	A3-	A4+	A4-	PL
R1	1	0	1	0	0	0	0	0	C1
R2	1	1	1	0	0	0	0	0	C2

2) Deep Learning Classification Method

The data were then split. Testing involved comparing the IndoBERT and LSTM methods. The model predictions resulted in unique powerset labels that represent several aspects and sentiments. The performance of the model was then evaluated using the testing data to assess the accuracy and effectiveness of the model developed for the M2 system.

All tools and libraries used in this study were implemented in the Python programming language. Libraries include Pandas, NumPy, Regular Expressions (re), NLTK, Sastrawi, nlp-id, scikit-learn, TensorFlow, Keras, Matplotlib, Seaborn, TQDM, Googletrans, OpenPyXL, Pickle, OS, AST, and IPython display.

III. RESULTS AND DISCUSSION

A. Test Scenarios for M1

Table V shows the test scenarios for M1. The test results showed that the global accuracy for the M1-1 test was 0.76. The M1-1 test was conducted using the best model for aspect classification with a tree count of 50 and a depth of 30, and sentiment classification with a tree count of 10 and a depth of 30. Sentiment classification A2 involved a tree count of 5 and a depth of 15. Sentiment classification A3 involved a tree count of 5 and a depth of 7. Finally, sentiment classification A4 involved a tree count of 50 and a depth of 30.

For the M1-2 test, the best accuracy achieved was 0.81. This test was performed with an embedding size of 128 and 20 epochs.

TABLE V. TESTING SCENARIO FOR M1

Scenario	Lexicon	OE	CM	Amount of trees	Depth	Embedding size
M1-1	TextBlob	Yes	RF	{5, 10, 50, 100. 120} [30]	{1,5,7,10, 15,25,30} [30]	-
M1-2	VADER	Yes	LSTM	-	-	{32, 64, 128, 256}[31]

OE = Opinion Extraction; CM = Classification Method; RF = Random Forest

B. Test Scenarios for M2

Table VI shows the test scenarios for M2. There were four types of scenarios with changes in epochs and the number of training data PLs. Figures 5 and 6 show the accuracy results for each scenario.

TABLE VI. TESTING SCENARIO FOR M2

Scenario	Imbalance method	PL	Epochs [31]	CM	amount of PL training
M2-1	No	Yes	{10,5,3}	IndoBERT	{20,15,10}
M2-2	No	Yes	{10,5,3}	LSTM	{20,15,10}
M2-3	ROS	Yes	{10,5,3}	IndoBERT	{20,15,10}
M2-4	ROS	Yes	{10,5,3}	LSTM	{20,15,10}

Figure 5 shows that the number of labels and the number of epochs affect the performance of the IndoBERT model. The best accuracy was 0.78, demonstrated by the M2 system with 10 PLs. Overall, these results confirm that without balancing, the model produces seemingly high accuracy but is not followed by even classification ability across classes. Meanwhile, the LSTM model was trained without any balancing techniques. The test results showed more stable performance compared to the Random Over-Sampling (ROS) scenario, with consistent accuracy in the range of 0.38–0.41 for configurations of 20, 15, and 10 PLs across all epoch variations (3, 5, and 10).

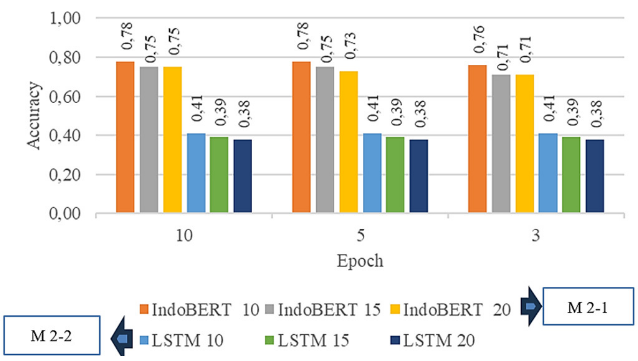

Fig. 5. M2-1 and M2-2 results.

In Figure 6, it can be seen that ROS improves accuracy only by 0.01. The evaluation results show that the best accuracy was obtained by the model with 10 PLs and 5 epochs at 0.79. However, this indicates that ROS is not fully capable of overcoming the challenge of multi-label classification with a complex number of classes, because this is only 10 labels out of the total 69 in the dataset. Meanwhile, for the LSTM model with the balanced dataset, the accuracy is not good. This indicates that the main challenge is not solely on data imbalance but on the complexity of the PL itself, which produces a large number of classes.

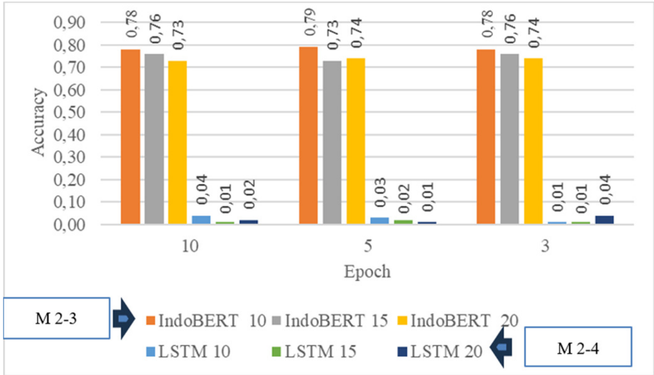

Fig. 6. Results of M2-3 and M2-4.

C. *Discussion*

The best-performing model may vary depending on the M-ABSA approach employed. In this study, the feature engineering-based approach with rule-based opinion extraction outperformed the label transformation approach. Several experimental scenarios were conducted using models M1 and M2, with the best achieving an accuracy of 0.81. Nevertheless, the proposed system can be used to support improvements in healthcare facility service management.

D. *Limitations*

The aspects used are still determined by the researcher and are not automatically extracted from the review data in the dataset. Handling imbalanced data in the M1 system dataset significantly impacts the accuracy of the system. The created rule base is a weakness in the M2 system. The data used is Indonesian, but social media reviews can be in non-standard, regional, or foreign languages. Further research is needed to address these limitations and improve performance.

IV. CONCLUSION

The test results showed that the application of feature engineering techniques using opinion extraction with DL (LSTM) achieved the best accuracy of 0.81. This technique handled all multi-label combinations effectively, but global accuracy was still influenced by the classification of the aspect and the sentiment classification per aspect. The rule base used for opinion analysis also presented a weakness in this system. Therefore, further development is needed regarding the formation of a rule base for review opinion extraction. The application of multiple methods to support the ultimate goal of multi-label aspect-based sentiment analysis prediction also impacted system error.

The application of PL to M-ABSA, although achieving an accuracy of 0.79, was deemed inadequate due to the primary problem of data imbalance. Therefore, special handling of this issue is needed by implementing a balancing system model for M-ABSA.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the results of this study.

ACKNOWLEDGMENT

The authors sincerely thank the University of Trunojoyo Madura (UTM) and the Indonesian Ministry of Higher Education, Research, and Technology (KEMDIKTISAINTEK) for the opportunity and support in conducting this research. This research is an output of a BIMA grant 2025 with contract numbers 120/C3/DT.05.00/PL/2025 and B/023/UN46.1/PT.01.03/BIMA/PL/2025.

DATA AVAILABILITY

The data that support the findings of this study will be made available on request.

REFERENCES

[1] S. Sapkota, M. R. Tiwary, and K. B. Zare, "Access to Healthcare Facilities and Social Well-being in Urban Areas," *Journal of Applied. Bioanalysis*, vol. 10, no. 2, pp. 243–251, 2024.

[2] Z. Gizaw, T. Astale, and G. M. Kassie, "What improves access to primary healthcare services in rural communities? A systematic review," *BMC Primary Care*, vol. 23, no. 1, Dec. 2022, Art. no. 313, https://doi.org/10.1186/s12875-022-01919-0.

[3] R. Irawan and E. Pasaribu, "The Effect Of Health Expenditure On Life Expectancy In Bengkulu," *Jurnal Ekonomi*, vol. 29, no. 3, pp. 552–569, Dec. 2024, https://doi.org/10.24912/je.v29i3.2609.

[4] A. I. A. Rahim, M. I. Ibrahim, K. I. Musa, S. L. Chua, and N. M. Yaacob, "Assessing Patient-Perceived Hospital Service Quality and Sentiment in Malaysian Public Hospitals Using Machine Learning and Facebook Reviews," *International Journal of Environmental Research and Public Health*, vol. 18, no. 18, Sept. 2021, https://doi.org/10.3390/ijerph18189912.

[5] L. Pokhrel, A. Kumar, P. Garg, N. Anand, and N. Singh, "AI and IoT in Global Health: Ethical Lessons From Pandemic Response," in *Development and Management of Eco-Conscious IoT Medical Devices*, IGI Global Scientific Publishing, 2026, pp. 367–394.

[6] A. T. Eshlaghy, A. Daneshvar, A. Peivandizadeh, A. R. S Senathirajah, and I. Ibrahim, "Designing a Sustainable Model for Providing Health Services Based on the Internet of Things and Meta-Heuristic Algorithms," *International Journal of Supply and Operations Management*, vol. 12, no. 1, pp. 28–47, Feb. 2025, https://doi.org/10.22034/ijsom.2023.110025.2827.

[7] Y. Hapsari and A. C. Sjaaf, "Effect of Hospital Accreditation on Patient Safety Culture and Satisfaction: A Systematic Review," *The International Conference on Public Health Proceedings*, Bangkok, Thailand, vol. 4, no. 02, pp. 547–555, https://doi.org/10.26911/the6thicph-FP.04.42.

[8] M. J. Alhawajreh, W. J. Jackson, and A. S. Paterson, "Healthcare professionals' perceptions about implementing accreditation as a strategy to improve healthcare quality and organisational performance: a cross-sectional survey study," *PLOS ONE*, vol. 20, no. 3, 2025, Art. no. e0320664, https://doi.org/10.1371/journal.pone.0320664.

[9] M. J. Alhawajreh, A. S. Paterson, and W. J. Jackson, "Impact of hospital accreditation on quality improvement in healthcare: A systematic review," *PLOS ONE*, vol. 18, no. 12, Dec. 2023, Art. no. e0294180, https://doi.org/10.1371/journal.pone.0294180.

[10] A. Vuohijoki, L. Ristolainen, J. Leppilahti, S. M. Kivivuori, and H. Hurri, "Impact of joint commission international accreditation on occupational health and patient safety: A systematic review," *PLOS One*, vol. 20, no. 6, June 2025, Art. no. e0325894, https://doi.org/10.1371/journal.pone.0325894.

[11] P. C. C. A. Nasution, D. Ayuningtyas, A. Bachtiar, and B. Besral, "The Development and Impact of Hospitals Accreditation in Southeast Asia: A Scoping Review," *Journal of Health Research*, vol. 39, no. 2, May 2025, https://doi.org/10.56808/2586-940X.1139.

[12] S. B. Syed, S. Leatherman, N. Mensah-Abrampah, M. Neilson, and E. Kelley, "Improving the quality of health care across the health system," *Bulletin of the World Health Organization*, vol. 96, no. 12, Dec. 2018, Art. no. 799, https://doi.org/10.2471/BLT.18.226266.

[13] C. Murray, L. Mitchell, J. Tuke, and M. Mackay, "Revealing patient-reported experiences in healthcare from social media using the design-acquire-process-model-analyse-visualise framework," *DIGITAL HEALTH*, vol. 10, Jan. 2024, Art. no. 20552076241251715, https://doi.org/10.1177/20552076241251715.

[14] M. I. Khan, Z. U. Rahman, M. A. Saleh, and S. U. Z. Khan, "Social Media and Social Support: A Framework for Patient Satisfaction in Healthcare," *Informatics*, vol. 9, no. 1, Mar. 2022, https://doi.org/10.3390/informatics9010022.

[15] E. I. Setiawan, "Aspect-Based Sentiment Analysis of Healthcare Reviews from Indonesian Hospitals based on Weighted Average Ensemble," *Journal of Applied Data Sciences*, vol. 5, no. 4, pp. 1579–1596, Dec. 2024, https://doi.org/10.47738/jads.v5i4.328.

[16] P. Bhatia and R. Nath, "Using Sentiment Analysis in Patient Satisfaction: A Survey," *Advances in Mathematics: Scientific Journal*, vol. 9, no. 6, pp. 3803–3812, July 2020, https://doi.org/10.37418/amsj.9.6.59.

[17] A. Siddiqua and H. C. Nagaraj, "Design and Development of an Efficient Ensemble Model for Aspect-Based Sentiment Analysis," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 24964–24969, Aug. 2025, https://doi.org/10.48084/etasr.11176.

[18] J. Li *et al.*, "Identifying healthcare needs with patient experience reviews using ChatGPT," *PLOS ONE*, vol. 20, no. 3, Mar. 2025, Art. no. e0313442, https://doi.org/10.1371/journal.pone.0313442.

[19] E. I. Setiawan, P. Tjendika, J. Santoso, F. X. Ferdinandus, G. Gunawan, and K. Fujisawa, "Aspect-Based Sentiment Analysis of Healthcare Reviews from Indonesian Hospitals based on Weighted Average Ensemble," *Journal of Applied Data Sciences*, vol. 5, no. 4, pp. 1579–1596, Oct. 2024, https://doi.org/10.47738/jads.v5i4.328.

[20] C. A. S. Araujo, M. M. Siqueira, and A. M. Malik, "Hospital accreditation impact on healthcare quality dimensions: a systematic review," *International Journal for Quality in Health Care*, vol. 32, no. 8, pp. 531–544, Nov. 2020, https://doi.org/10.1093/intqhc/mzaa090.

[21] K. Lewis and R. Hinchcliff, "Hospital accreditation: an umbrella review," *International Journal for Quality in Health Care*, vol. 35, no. 1, Feb. 2023, Art. no. mzad007, https://doi.org/10.1093/intqhc/mzad007.

[22] A. Maretta and A. Meiriza, "Aspect-Based Sentiment Analysis of Hospital Service Reviews Using Fine-Tuned IndoBERT," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2541–2551, Oct. 2025, https://doi.org/10.30871/jaic.v9i5.10765.

[23] N. C. Mei, S. Tiun, and G. Sastria, "Multi-Label Aspect-Sentiment Classification on Indonesian Cosmetic Product Reviews with IndoBERT Model," *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 11, Nov. 2024, Art. no. 712, https://doi.org/10.14569/ijacsa.2024.0151168.

[24] F. M. Alotaibi, "A Machine-Learning-Inspired Opinion Extraction Mechanism for Classifying Customer Reviews on Social Media," *Applied Sciences*, vol. 13, no. 12, June 2023, Art. no. 7266, https://doi.org/10.3390/app13127266.

[25] D. Kumar and F. Ahamad, "A Review on Challenges in Recent Opinion Extraction Techniques," in *International Journal of Innovative Research in Computer Science and Technology (IJIRCST)*, Mar. 2024, pp. 13–17, https://doi.org/10.55524/CSISTW.2024.12.1.3.

[26] L. W. Ku, Y. T. Liang, and H. H. Chen, "Opinion extraction, summarization and tracking in news and blog corpora," in *Papers from the 2006 AAAI Spring Symposium*, 2006.

[27] A. Chiche and B. Yitagesu, "Part of speech tagging: a systematic review of deep learning and machine learning approaches," *Journal of Big Data*, vol. 9, no. 1, Jan. 2022, Art. no. 10, https://doi.org/10.1186/s40537-022-00561-y.

[28] M. Raees and S. Fazilat, "Lexicon-Based Sentiment Analysis on Text Polarities with Evaluation of Classification Models." arXiv, 2024, https://doi.org/10.48550/ARXIV.2409.12840.

[29] E. C. Narendra, A. A. Arifiyanti, and T. L. I. Sugata, "Enhancing Aspect-Based Sentiment Analysis in Imbalanced Multilabel Datasets using Resampling and Classifiers for Digital Signature Applications," *Aviation Electronics, Information Technology, Telecommunications, Electricals, and Controls (AVITEC)*, vol. 7, no. 2, pp. 195–207, June 2025, https://doi.org/10.28989/avitec.v7i2.3023.

[30] A. M. Shetty, M. F. Aljunid, D. H. Manjaiah, and A. M. S. S. Afzal, "Hyperparameter Optimization of Machine Learning Models Using Grid Search for Amazon Review Sentiment Analysis," in *Data Science and Applications*, 2024, pp. 451–474, https://doi.org/10.1007/978-981-99-7814-4_36.

[31] E. Ben-David, C. Rabinovitz, and R. Reichart, "PERL: Pivot-based Domain Adaptation for Pre-trained Deep Contextualized Embedding Models," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 504–521, July 2020, https://doi.org/10.1162/tacl_a_00328.