

# Deep Learning Approach for Aspect-Based Sentiment Analysis on Indonesian Hospitals Reviews

1<sup>st</sup> Muhamad Fahmi

Master Program in Artificial  
Intelligence, Department of Computer  
Science and Electronics, Faculty of  
Mathematics and Natural Sciences  
Universitas Gadjah Mada, Indonesia  
Yogyakarta, Indonesia  
muhamadfahmi@mail.ugm.ac.id

2<sup>nd</sup> Faturahman Yudianto

Master Program in Artificial  
Intelligence, Department of Computer  
Science and Electronics, Faculty of  
Mathematics and Natural Sciences  
Universitas Gadjah Mada, Indonesia  
Yogyakarta, Indonesia  
f.yudianto@mail.ugm.ac.id

3<sup>rd</sup> Naurah Nazhifah

Master Program in Artificial  
Intelligence, Department of Computer  
Science and Electronics, Faculty of  
Mathematics and Natural Sciences  
Universitas Gadjah Mada, Indonesia  
Yogyakarta, Indonesia  
naurnazhifah1399@mail.ugm.ac.id

4<sup>th</sup> Yunita Sari

Department of Computer Science and  
Electronics, Faculty of Mathematics  
and Natural Sciences  
Universitas Gadjah Mada, Indonesia  
Yogyakarta, Indonesia  
yunita.sari@ugm.ac.id

5<sup>th</sup> Afiahayati

Department of Computer Science and  
Electronics, Faculty of Mathematics  
and Natural Sciences  
Universitas Gadjah Mada, Indonesia  
Yogyakarta, Indonesia  
afia@ugm.ac.id

**Abstract**—Hospitals are individual health service facilities that provide inpatient and outpatient care, therefore, quality hospital services, facilities, and resources are a necessity and must be met by every hospital. To assess the quality of hospital services, facilities, and resources, we can see reviews from users through reviews available on Google Maps. Using natural language understanding, we can identify the sentiment associated with each particular aspect. In this research, we developed a deep learning model which can classify three aspects (services, human resources, and facilities) with four sentiments (positive, negative, neutral, and none) using a multiclass-multioutput deep learning model, ensuring both aspect and sentiment classification within a single model. Our research concludes that the BERT-LSTM model demonstrates promising performance with an F1-score of 0.597, but faces challenges in accurately classifying neutral sentiment across all aspects due to the lack of neutral sentiment labels.

**Keywords**—*aspect based sentiment analysis, hospital, deep learning, BERT, LSTM, BiLSTM*

## I. INTRODUCTION

Healthcare institutions, such as hospitals, that are directly connected to patients should give precedence to ensuring safety, quality, non-discrimination, and the provision of effective health services. This involves prioritizing patient interests in alignment with established hospital service standards. Patients as the main users of health services have the right to obtain security and safety during treatment in hospitals[1]. The quality of hospital service can be identified through its review from the internet in a form of user-generated content.

User generated content which is often used to assess the quality of health services is usually presented in the form of ratings and reviews. The reviews are considered to be able to provide indirect experiences to other customers. Apart from the customer side, reviews in user generated content also provide benefits from the marketing business side [2]. Positive reviews are deemed to indirectly contribute to promotion,

especially when we are aware of the characteristics associated with each sentiment.

Positive reviews shared by customers can strengthen potential customers' positive impressions of a place, encouraging their interest in visiting that place after reading positive testimonials from other customers. On the other hand, negative reviews may be a factor that potential customers consider when planning a visit to the establishment. However, with the large number of reviews available, potential consumers often only read the reviews briefly, which are generally located at the top, resulting in difficulties in forming conclusions and making decisions. [3]. Aspect-based sentiment analysis is a method for determine the tendency of sentiment polarity in an opinion sentence, by referring to certain aspects that have been previously identified. To get sentiment insight in each review, this research proposes to use aspect-level sentiment analysis.

Sentiment analysis is one of the essential topics in natural language processing (NLP). This process aims to identify and extract subjective information, and then categorize it into three main polarities: positive, negative, and neutral. Sentiment analysis techniques based on machine learning use labelled data to train models that identify text as positive, negative, or neutral [4]. However, most conventional sentiment analysis is only able to recognize one polarity for each sentence without considering the possibility of more than one sentiment in one sentence. Therefore, a sentiment analysis approach is needed that focuses on certain aspects to be able to understand and classify the various sentiment polarities that may arise in a context more accurately.

In this study, an aspect-based sentiment analysis was conducted to determine the sentiment of hospital visitors on the aspects of service, hospital facilities, and human resources. And each aspect is rated by four sentiments, namely positive, negative, neutral, and none.

## II. RELATED WORKS

Aspect-based sentiment analysis (ABSA) is a type of sentiment analysis that allows research of all sentiments related to each aspect. There are various methods that can be used to carry out sentiment analysis based on this aspect. In research [5] [6] [7], the researcher utilizes the aspect extraction method with POS Tagger to identify aspects and opinion words before classifying aspects and sentiments. Sentiment classification in Aspect Based Sentiment Analysis (ABSA) can be done through various methods. Available options include performing sentiment classification for all aspects or each aspect [8] [9].

Research by Cahyaningtyas et.al. [10] proposed to create ABSA with eight deep learning methods namely Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional-LSTM (BiLSTM), Attention BiLSTM, Convolutional Neural Network (CNN), CNN-LSTM, and CNN-BiLSTM, on 6 class aspects. In that case, the aspect classification model and sentiment classification model were built separately. The results showed that LSTM got the best model for aspect classification and for sentiment classification, CNN model got the best model.

ABSA can be a challenging task as an opinion can contain multiple aspects [11] [12] [13] [14]. Bidirectional Encoder Representation from Transformer (BERT) models are widely used for ABSA [15] [16] [17] [18]. BERT explores the recent bidirectional encoder representation of Transformer, which processes text by considering the context of a word from both the left and right sides [19]. As such, BERT can produce a more semantic representation of text, where each word is converted into an embedding vector that depends on its context in a sentence. Additionally, BERT is a language model that has been pre-trained on large text corpora and can be adapted for supervised derivation tasks, such as Aspect-Based Sentiment Analysis (ASBA). In fact, most of the current research in ASBA is exploring various customization strategies that can be applied to BERT.

## III. METHODOLOGY

### A. Research Methodology

#### 1. Data Collection

Data collection is the process of retrieving information or data from various sources for a specific purpose. In this study, data collection is carried out using a scraping technique that uses the Selenium library to retrieve reviews from each hospital, especially those located on Sumatra Island, Kalimantan Island, and Yogyakarta Province. The reviews are taken from the hospital Google Maps review page. The scraped review texts are then stored in a text file. The sample of scraping results can be seen in Figure 1.

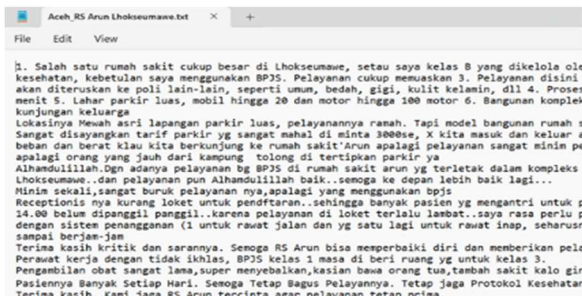


Fig. 1. Data Scraping Result

#### 2. Data Labeling

The results of the scraping process are then labeled using a data labeling toolkit namely Label Studio. This labeling process focuses on three aspects, namely hospital services, human resources, and facilities. Each aspect was then rated from four labels. There are three labels that show the sentiment polarity, namely positive, negative, and neutral, also one label, the "none" label, for non-existent sentiment labels in a certain aspect on the text data. The labeling process can be seen in Figure 4.

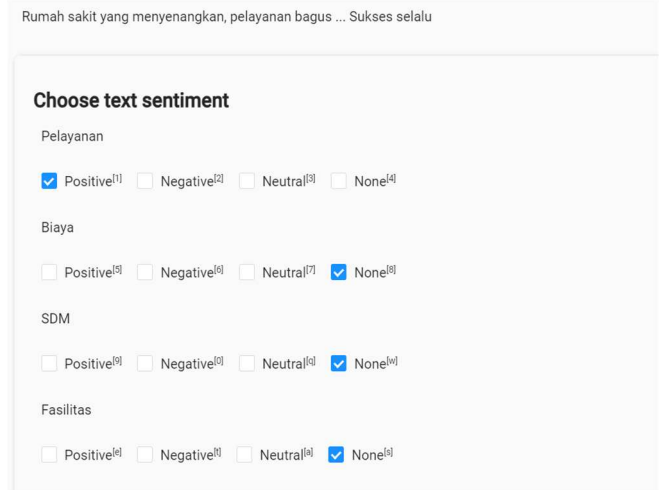


Fig. 2. Data Labeling

#### 3. Data Splitting

The labeled dataset is divided into two distinct sets: the training set and the test set. The training set contains 1841 reviews while the test set contains 461 reviews. Notably, the validation set is drawn from the training set, comprising 20% of its proportion.

#### 4. Preprocessing

In this research, the preprocessing stage is carried out in four stages, namely case folding, removal of numbers and punctuation marks, stopword removal, and tokenization. Case folding is a text processing technique used to convert all letters in a text to lowercase. The purpose is to eliminate the differences that occur in the same words due to the difference in upper and lower case letters.

After case folding, number and punctuation removal is done to clean the text from unnecessary characters. That way, the text can be processed more easily and effectively because it only consists of relevant words and has a uniform form. Stopword removal was then implemented to remove words with no polarity.

Tokenization is the process of breaking down a text or document into smaller units, referred to as tokens. The tokenization process in this research is also carried out to form a sequence vector from the text containing a set of word ids and their padding values with a maximum vector length of 256 tokens. The tokenizer used for the tokenization process is BertTokenizer with vocabulary taken from the "indobert-base-p2" [20] model.

#### 5. Aspects + Sentiment Classification

There are three aspects, namely (1) services, (2) human resources, and (3) facilities. Each aspect is rated by four sentiments, namely positive, negative, neutral, and none. We use a multiclass-multioutput classification approach in the

deep learning model so both aspect and sentiment of the text can be classified in one model. Further explanation about the model is provided in the modeling section.

### B. Dataset

The used dataset in this research is in the form of reviews from each hospital, especially Sumatra Island, Kalimantan Island and Yogyakarta Province. The visualization of the number of review text for each aspect used can be seen in Figure 3.

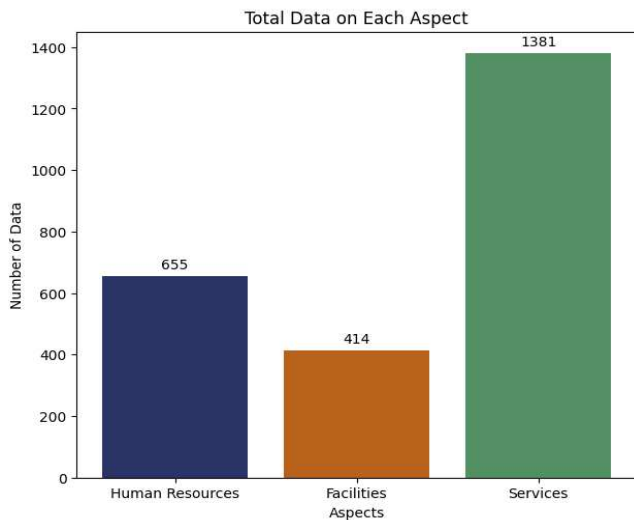


Fig. 3. Total Data for Each Aspect

The reviews mostly talk about the hospital service aspect with a total of 1381 data while the reviews which contain facilities and human resources aspect only numbered 414 and 655 respectively.

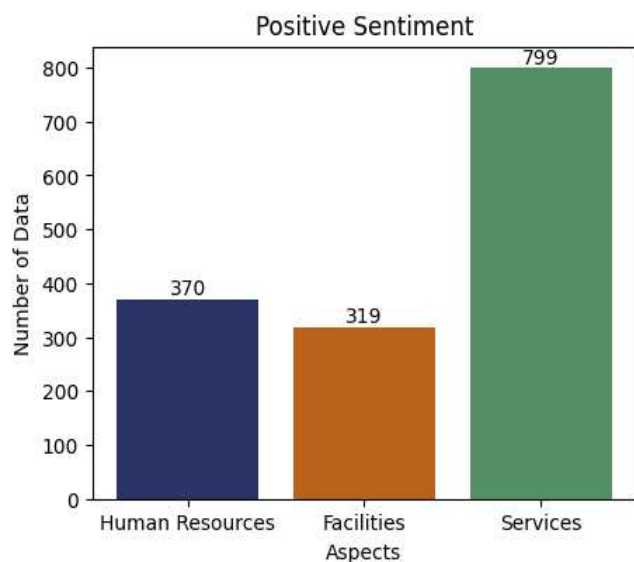


Fig. 4. Amount of Data for Positive Sentiment

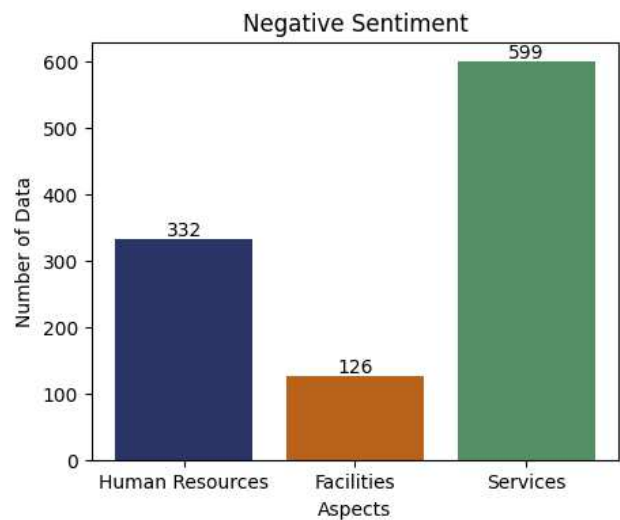


Fig. 5. Amount of Data for Negative Sentiment

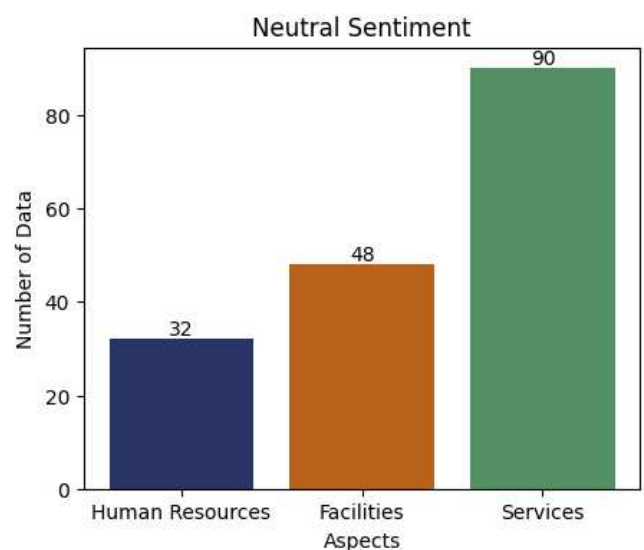


Fig. 6. Amount of Data for Neutral Sentiment

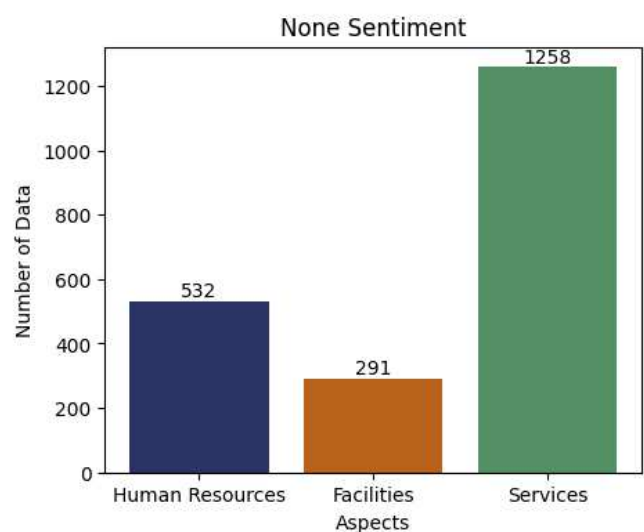


Fig. 7. Amount of Data for None Sentiment

The bar plots above explains the dataset that has been divided based on three aspects, with each aspect grouped into

four sentiment categories, namely positive, negative, neutral and none. Positive sentiment on the aspect of human resources has a total of 370, facilities amount 319, and services have the highest number which is 799. While negative sentiment on the aspect of human resources amounts to 332, facilities 126, and services reach 599. Neutral sentiment on the aspect of human resources amounts 32, facilities amount 48, and services amount 90. Finally, in the none sentiment category, the aspect of human resources touches 532, facilities are 291 and services have a total 1258.

From the bar plots, it can be seen that the service aspect has a very large number in each of the positive, negative, neutral, and none sentiment categories.

### C. Modeling

In this research, we use a multiclass - multioutput classification approach for aspect and sentiment classification. This approach allows the classifier to generate sentiment output on each aspect. The deep learning architecture used for multiclass-multioutput text classification generally consists of an embedding layer to represent the text, feature extraction layer using either RNN-base model such as LSTM or BERT model, and feedforward layers for the classification task. The number of feedforward layers is adjusted to the number of aspects to be classified, which is three aspects. Each feedforward layer has four nodes in the output layer according to the number of sentiment classes in each aspect. The architecture diagram of the model can be seen in Figure 8.

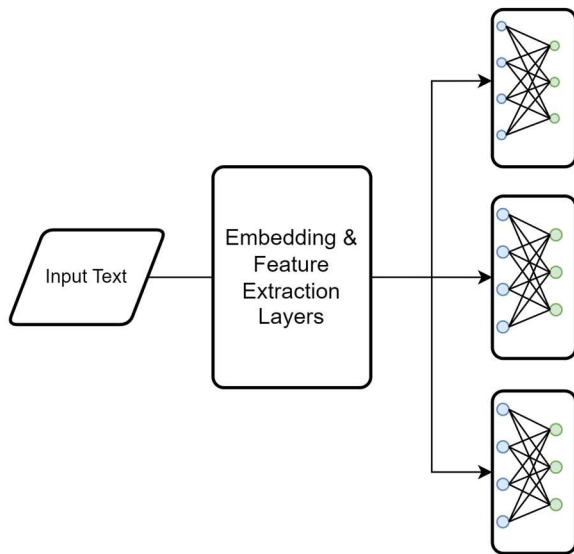


Fig. 8. Deep Learning Architecture for ABSA

In this study, we compare three deep learning architectures for the case of aspect-based sentiment analysis mainly in the embedding layer and feature extraction layer, namely Bidirectional-LSTM, BERT, and a combination of BERT and LSTM (BERT-LSTM). The three architectures are trained with the same input configuration. We set the maximum length of input token to 256 tokens and collate the dataset into batches with the batch size of 16.

#### 1. BERT

In this experiment, we used a pre-trained BERT model in Indonesian: IndoBERT [18]. The model we used is "indobert-base-p2". In this experiment, we also applied different learning rates of  $1.10^{-5}$  in the BERT layer and  $1.10^{-3}$  in the

feedforward layers. The architecture diagram is depicted in the following diagram.

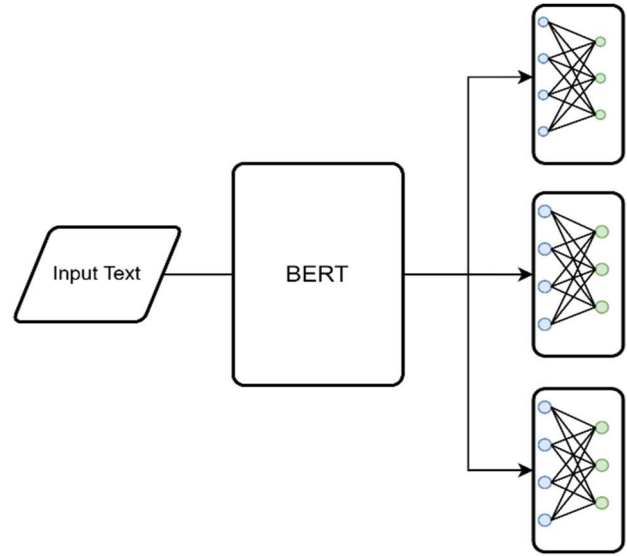


Fig. 9. BERT Architecture

The model was trained with the parameters shown in the following table.

TABLE I. BERT PARAMETERS

Parameters	
<i>optimizer</i>	AdamW
<i>learning rate on BERT</i>	$1e-5$
<i>learning rate on feed forward layers</i>	$1e-3$

#### 2. Bi-LSTM

In this experiment, we use the Bidirectional LSTM (Bi-LSTM) architecture in the feature extraction layer. Before the Bi-LSTM layer, an embedding layer is added provided by the Pytorch library which is a simple lookup table that stores the embedding vectors of the vocabulary. The architecture diagram is depicted in Figure 10.

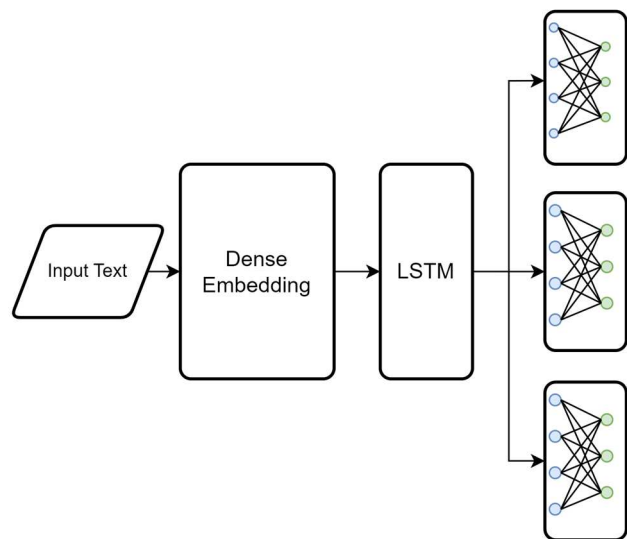


Fig. 10. Bi-LSTM Architecture

The model was trained with the parameters shown in table 3 below.

TABLE II. BI-LSTM PARAMETERS

Parameters	
<i>optimizer</i>	Adam
<i>learning rate</i>	1e-3
<i>number of layers</i>	5
<i>hidden dimension</i>	512
<i>embedding dimension</i>	768

### 3. BERT-LSTM

In this experiment we combine the BERT layer from the first experiment and the LSTM layer. In this experiment we also applied different learning rates of  $1.10^{-5}$  in the BERT layer and  $1.10^{-3}$  in the LSTM and feedforward layers. The architecture diagram is depicted in Figure 11.

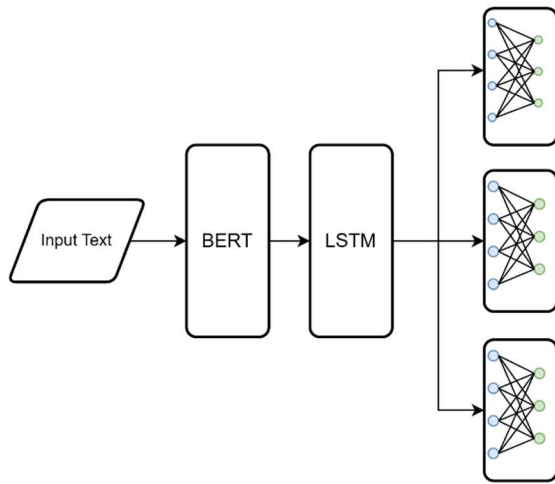


Fig. 11. BERT-LSTM Architecture

The model was trained with the parameters shown in table 4 below.

TABLE III. BERT-LSTM PARAMETERS

Parameters	
<i>optimizer</i>	AdamW
<i>learning rate on BERT</i>	1e-5
<i>learning rate on LSTM + feed forward layers</i>	1e-3
<i>number of layers on LSTM</i>	2
<i>hidden dimension on LSTM</i>	512

## IV. RESULT

Model evaluation is done by measuring the value of the macro average F1-score in each sentiment class in each aspect. The macro average F1-score value is used considering the unbalanced number of data. The results of the macro average F1-score value can be seen in the following Table.

TABLE IV. BI-LSTM PARAMETERS

		Bi-LSTM	BERT	BERT-LSTM
Services (Pelayanan)	Positive	0.80	0.85	0.86
	Neutral	0.00	0.23	0.08
	Negative	0.69	0.82	0.81
	None	0.68	0.71	0.73
	Average	0.54	<b>0.65</b>	0.62
Human Resources (Sumber Daya Manusia)	Positive	0.65	0.68	0.72
	Neutral	0.00	0.00	0.00
	Negative	0.52	0.66	0.61
	None	0.81	0.84	0.84
	Average	0.49	<b>0.54</b>	<b>0.54</b>
Facilities (Fasilitas)	Positive	0.55	0.67	0.68
	Neutral	0.14	0.22	0.37
	Negative	0.31	0.56	0.58
	None	0.89	0.92	0.91
	Average	0.47	0.60	<b>0.63</b>
Overall Average		0.5	<b>0.597</b>	<b>0.597</b>

From the results, it can be concluded that both the BERT and BERT-LSTM model achieve the best overall F1-score in the value of 0.597. Those models also achieve the same performance in sentiment classification on the human resources aspect. In the services aspect the BERT model achieve the macro average highest F1 score with 0.65 while in the facilities aspect BERT-LSTM achieve the highest macro average F1 score with 0.63. In terms of the model performance based on each aspect, the model performance on the human resources aspect still remains low due to the low number of training data that contains the human resources aspect. In overall aspect, neutral sentiment remains a challenge in all aspects and all models, likely due to a lack of labeled data. To address this limitation, enhancing the model's performance can be achieved through methods such as incorporating more training data, or consolidating neutral and non-existent sentiment labels. Data augmentation is also a potential solution to improve and expand the dataset, considering that the available data for model training is still limited and imbalanced.

## V. CONCLUSION

In this study, the BERT-LSTM model achieved the best overall F1-score in the value of 0.597. However, there was difficulty in classifying neutral sentiment in all aspects. This challenge is attributed to the lack of neutral sentiment labels in the dataset. To enhance the model's performance in recognizing neutral sentiment, steps such as adding training data or merging neutral sentiment labels with non-existent sentiment labels can be taken.

The results of the evaluation indicate significant potential in the model, particularly in the service aspect. Nevertheless, the primary challenge lies in improving the model's ability to classify neutral sentiment accurately across all aspects. Therefore, the primary conclusion is that the model holds



great promise, and there are opportunities for improvement to enhance its capability in classifying neutral sentiment across diverse contexts.

## REFERENCES

- [1] "PP\_Nomor\_47\_Tahun\_2021.pdf." Accessed: Nov. 16, 2023. [Online]. Available: [https://jdih.setkab.go.id/PUUdoc/176340/PP\\_Nomor\\_47\\_Tahun\\_2021.pdf](https://jdih.setkab.go.id/PUUdoc/176340/PP_Nomor_47_Tahun_2021.pdf)
- [2] *Integrated Uncertainty in Knowledge Modelling and Decision Making*, H. Seki, C. H. Nguyen, V.-N. Huynh, and M. Inuiguchi, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 393–405. doi: 10.1007/978-3-030-14815-7\_33.
- [3] P. R. Amalia and E. Winarko, "Aspect-Based Sentiment Analysis on Indonesian Restaurant Review Using a Combination of Convolutional Neural Network and Contextualized Word Embedding," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 3, Art. no. 3, Jul. 2021, doi: 10.22146/ijccs.67306.
- [4] M. Fahmi, S. Hidayat, and A. F. Hidayatullah, "APPLICATION OF LEXICON BASED FOR SENTIMENT ANALYSIS OF COVID-19 BOOSTER VACCINATIONS ON TWITTER SOCIAL MEDIA USING NAÏVE BAYES METHOD," *Jurnal Teknik Informatika (Jutif)*, vol. 3, pp. 1119–1124, Aug. 2022, doi: 10.20884/1.jutif.2022.3.4.565.
- [5] W.-H. Khong, L.-K. Soon, H.-N. Goh, and S.-C. Haw, "Leveraging Part-of-Speech Tagging for Sentiment Analysis in Short Texts and Regular Texts," in *Semantic Technology*, R. Ichise, F. Lecue, T. Kawamura, D. Zhao, S. Muggleton, and K. Kozaki, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 182–197. doi: 10.1007/978-3-030-04284-4\_13.
- [6] M. Afzaal, M. Usman, and A. Fong, "Predictive aspect-based sentiment classification of online tourist reviews," *Journal of Information Science*, vol. 45, no. 3, pp. 341–363, Jun. 2019, doi: 10.1177/0165551518789872.
- [7] B. Jang, M. Kim, G. Harerimana, S. Kang, and J. W. Kim, "Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism," *Applied Sciences*, vol. 10, no. 17, Art. no. 17, Jan. 2020, doi: 10.3390/app10175841.
- [8] M. Al-Smadi, B. Talafha, M. Al-Ayyoub, and Y. Jararweh, "Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews," *Int. J. Mach. Learn. & Cyber.*, vol. 10, no. 8, pp. 2163–2175, Aug. 2019, doi: 10.1007/s13042-018-0799-4.
- [9] D. Ekawati and M. L. Khodra, "Aspect-based sentiment analysis for Indonesian restaurant reviews," in *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*, Aug. 2017, pp. 1–6. doi: 10.1109/ICAICTA.2017.8090963.
- [10] S. Cahyaningtyas, D. H. Fudholi, and A. F. Hidayatullah, "Deep Learning for Aspect-Based Sentiment Analysis on Indonesian Hotels Reviews," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, Aug. 2021, doi: 10.22219/kinetik.v6i3.1300.
- [11] L. Zhang, "The Combined Application Technology of the Bank's TwoDimensional Aggregate Payment Code and the Vegetable Field Traceability System," *Int. J. Informatics Inf. Syst.*, vol. 5, no. 4, pp. 150–155, Dec. 2022.
- [12] B. N. D. Santos, R. M. Marcacini, and S. O. Rezende, "Multi-Domain Aspect Extraction Using Bidirectional Encoder Representations From Transformers," *IEEE Access*, vol. 9, pp. 91604–91613, 2021, doi: 10.1109/ACCESS.2021.3089099.
- [13] D. Arbian Sulistyo et al., "LSTM-Based Machine Translation for Madurese-Indonesian," *J. Appl. Data Sci.*, vol. 4, no. 3, pp. 189–199, Sep. 2023.
- [14] A. Wang, Z. Qin, and Y.-H. Dong, "Development of an IoT-Based Parking Space Management System Design," *Int. J. Appl. Inf. Manag.*, vol. 3, no. 2, pp. 91–100, Jul. 2023.
- [15] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "Attentional Encoder Network for Targeted Sentiment Classification," vol. 11730, 2019, pp. 93–103. doi: 10.1007/978-3-030-30490-4\_9.
- [16] B. Zeng, H. Yang, R. Xu, W. Zhou, and X. Han, "LCF: A Local Context Focus Mechanism for Aspect-Based Sentiment Classification," *Applied Sciences*, vol. 9, no. 16, Art. no. 16, Jan. 2019, doi: 10.3390/app9163389.
- [17] A. Rietzler, S. Stabinger, P. Opitz, and S. Engl, "Adapt or Get Left Behind: Domain Adaptation through BERT Language Model Finetuning for Aspect-Target Sentiment Classification." arXiv, Nov. 19, 2019, doi: 10.48550/arXiv.1908.11860.
- [18] A. Karimi, L. Rossi, and A. Prati, "Adversarial Training for Aspect-Based Sentiment Analysis with BERT." arXiv, Oct. 23, 2020. doi: 10.48550/arXiv.2001.11316.
- [19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.
- [20] B. Wilie et al., "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, Suzhou, China: Association for Computational Linguistics, Dec. 2020, pp. 843–857. Accessed: Dec. 16, 2022. [Online]. Available: <https://aclanthology.org/2020.aacl-main.85>