

Aspect-Based Sentiment Analysis for Indonesian Tourist Attraction Reviews Using Bidirectional Long Short-Term Memory

Dwi Intan Af'idah¹, Puput Dewi Anggraeni², Muhammad Rizki³, Aji Bagus Setiawan⁴, Sharfina Febbi Handayani⁵

^{1,3,4,5}*Informatics Engineering, ³Hospitality Management - Polytechnic of Harapan Bersama, Indonesia*

¹dwiintanafidah@poltektegal.ac.id, ²puput.dewi@poltektegal.ac.id,

³19090055.rizki@student.poltektegal.ac.id,

⁴19090061.setiawan@student.poltektegal.ac.id, ⁵sharfina.handayani@poltektegal.ac.id

Abstract - The tourism sector in Indonesia experienced growth and made a positive contribution to the national economy, but this growth has yet to reach its target. Therefore, the government of Indonesia has implemented a sustainable tourism development program by establishing ten priority tourism destinations. Aspect-based sentiment analysis (ABSA) towards tourist attraction reviews can assist the government in developing potential goals. The ABSA process compares with two deep learning models (LSTM and Bi-LSTM), which are considered to obtain good performance in text analysis. The shortcomings of previous ABSA research should have examined the performance of the aspect classification and sentiment classification models sequentially. This makes the performance obtained from the ABSA task invalid. Thus, this study is conducted to determine the version of the aspect classification model and the sentiment classification model individually and simultaneously. This study aims to develop an aspect-based tourist attraction sentiment analysis as an intelligent system solution for sustainable tourism development by applying the binary relevance mechanism and the best deep learning model from LSTM or Bi-LSTM. The test results showed that Bi-LSTM was superior in aspect and sentiment classification individually and simultaneously. Likewise, the aspect classification and sentiment classification test results sequentially Bi-LSTM outperformed that of LSTM. The average accuracy and f1 score of Bi-LSTM are 92.22% and 71.06%. Meanwhile, LSTM obtained 90.63% of average precision and 70.4% of f1 score.

Keywords: Aspects-based sentiment analysis, LSTM, Bi-LSTM, tourism

I. INTRODUCTION

Indonesia's tourism sector is experiencing growth and positively contributing to the national economy. But, according to the Travel and Tourism Competitiveness Index (TTCI) ranking, Indonesia was ranked 40th in

2019, which still lags behind Singapore, Malaysia, and Thailand [1]. In comparison, tourism is now a global industry that involves millions of people in international and domestic travel every year. Therefore, tourism development strategies are essential to pay attention to in building sustainable tourism [2].

The Government of Indonesia has implemented a sustainable tourism development program by establishing ten priority tourism destinations since 2016. However, in their development, ten priority tourist destinations need help [3]. The issue of this tourist attraction needs to be monitored by the government and stakeholders of the tourism industry so that the problematic aspects of a tourist attraction get special attention. Information related to the issue of tourist attractions can be obtained from tourist reviews on social media. Therefore, an aspect-based sentiment analysis application is needed to summarise visitor reviews based on the aspect categories and then conduct sentiment analysis with the result that a percentage of positive or negative of each specified aspect is obtained [4].

Aspect-Based Sentiment Analysis (ABSA) is a process of extracting the polarity of opinion according to the specific aspect. There are two main processes in ABSA, such as the aspect extraction and the classification of sentiment aspect [5].

ABSA commits the analysis in several procedures. Aspect extraction using contextual features is implemented in research [6]-[7]. Studies of [8]-[9] proposed Conditional Random Field (CRF) models to obtain the best sequence of POS taggers in ABSA research. While in research [10]-[11], CRF is used to aspect opinion target expressions (OTEs) extraction before sentiment classification is done. Furthermore, the study researched the performance of several dependency relations of aspect extraction with POS tag patterns in

ABSA [12]. Another feature extraction, ABSA, also can be solved by the binary relevance strategy. The binary relevance strategy is suitable for transforming multi-label problem classification into binary classification. The binary relevance technique builds models for as many classes as aspects, and each aspect acquires its classification model [13]-[14].

Research of ABSA is generally solved using machine learning and deep learning. There are several experimentations of ABSA which use a Support Vector Machine (SVM) [15]-[17]. Meanwhile, the study [18] explained the various machine learning techniques on ABSA, which use numerous domain data. The development of machine learning that produces deep learning has also become a concern in the ABSA field. Further, deep learning is currently adopted to handle ABSA, such as in research [19]-[20]. LSTM (Long Short-Term Memory) as a deep learning model proves that it obtains good performance in text data [21]. LSTM was further developed into Bidirectional Long-Short Term Memory (Bi-LSTM). The study of text data analysis of various languages showed that Bi-LSTM offers better accuracy than the usual LSTM method [22]-[24].

Several studies regarding the ABSA system for the Indonesian language with numerous target domains have been executed recently. Reviews of Indonesian restaurants as datasets are used in studies [8] and [25]. While in studies [26], [14], [27], the ABSA method handles datasets in the form of Indonesian marketplace reviews. Furthermore, there are several target domains in the Indonesian language, namely hotel reviews [14], digital wallet reviews [28], and tweets about Indonesian presidential elections [17]. Moreover, studies [29]-[30] adopted the ABSA method for Indonesian tourism destination reviews.

Previous research regarding ABSA for tourism destination review [29]-[30] did not use reviews of 10 priority destinations in the Indonesian government program. In addition, research [30] entirely adopted machine learning for ABSA. LSTM and Bi-LSTM as deep learning provide better performance than classical machine learning in text analysis. Hence, comparing LSTM and Bi-LSTM is essential to examine in ABSA research. Another area for improvement is the shortcomings of previous ABSA research [13]-[14] that should have read the performance of the aspect classification and sentiment classification models sequentially. Still, it only individually examined the performance of the aspect and sentiment classification models. While in the ABSA process with the relevance strategy mechanism, a document requires aspect and

sentiment analysis simultaneously, which is not separate and stand-alone. In previous research [13]-[14], the evaluation of models was measured through accuracy, whereas dataset imbalance often occurs in ABSA cases. Evaluation models based on the F1 score should be done on datasets in the ABSA process that often need to be more balanced. So, a specific solution is required to develop an aspect-based sentiment analysis model for ten priority tourist destinations using LSTM and Bi-LSTM with a binary relevance strategy mechanism where the evaluation of the resulting model focuses on the f1 score. Meanwhile, model evaluation is conducted to determine the performance of the aspect classification model and the sentiment classification model individually and simultaneously.

This study aims to generate the best deep learning models for ten priority tourist destination reviews using a binary relevance strategy to solve multi-label classification. In this study, ABSA using a binary relevance strategy will develop four models of aspect classification: attraction, accessibility, facility, and accommodation. This study likewise produces a sentiment classification model with two positive and negative polarities for every aspect. In addition, this research will also show the performance comparison of LSTM and Bi-LSTM on the ABSA process used for aspect classification, for sentiment classification of each aspect, as well as the performance of aspect classification and sentiment classification sequentially, which previous research has yet to do.

The contributions of this study are to produce a labelled corpus based on aspect category (attraction, accessibility, facility, and accommodation) and sentiment classification (positive or negative) of tourist destination reviews in the Indonesian language also reveals that the Bi-LSTM method performs better than LSTM for ABSA of tourist destination reviews in the Indonesian language. This research produces the eight best Bi-LSTM models of aspect and sentiment classification. Further, the eight best Bi-LSTM were obtained to classify the aspect categories and sentiments. In addition, this study also executes aspect classification and sentiment classification in sequence, resulting in the performance of aspect classification and sentiment classification models in sequence, which has yet to be done in previous research.

II. METHOD

To analyse ABSA tasks, this research suggests two primary approaches using deep learning methods: (i) a long short-term memory method (LSTM) and (ii) a bidirectional long short-term memory method

(BiLSTM). Before the training process, the data is pre-trained first using word2vec. At the same time, the dataset used in this research is built by scraping and manually labelling data by experts. This research comprises six stages shown in Fig. 1.

A. Data Collection

Data collection in this study is done by collecting reviews of 10 priority attractions gathered from the Tripadvisor website using the web scraping method [31]. The ten priority tourist attractions are Lake Toba, Mandalika, Morotai, Tanjung Lesung, Labuan Bajo, Thousand Island, Wakatobi, Tanjung Kelayang, Bromo Tengger Semeru, and Borobudur [3]. Web Scraping is downloading documents from full web pages for more specific data retrieval or taking certain parts of a website. Knowledge of a website's HTML (Hypertext Markup Language) elements is needed to obtain the appropriate data. Automatic web scraping can automatically retrieve data from the website to speed up data collection. This scraping process uses Python libraries called BeautifulSoup [31].

B. Pre-processing

Text pre-processing aims to clean and tidy up the data, making it easy to use in the following process. The following are the stages of pre-processing. The first, case folding, converts all the letters' characters in a sentence to lowercase. The second filtering eliminates illegal document characters such as punctuation, symbols, numbers, etc. The third, tokenisation, is breaking a text document into tokens word-for-word. The fourth, Slang word conversion is the process of changing non-standard words into common words, is a stage of slang word conversion. The fifth, Stop-word removal, is taking important words and removing words considered unimportant. Stop-word removal aims to eliminate words that often appear but have no contribution to the data analysis [32]. The last, stemming, is mapping and decomposing a word's form into its base word form. The Nazief and Andriani algorithms are used in the stemming process, a special stemming algorithm for Indonesian text with a better accuracy percentage than others [33].

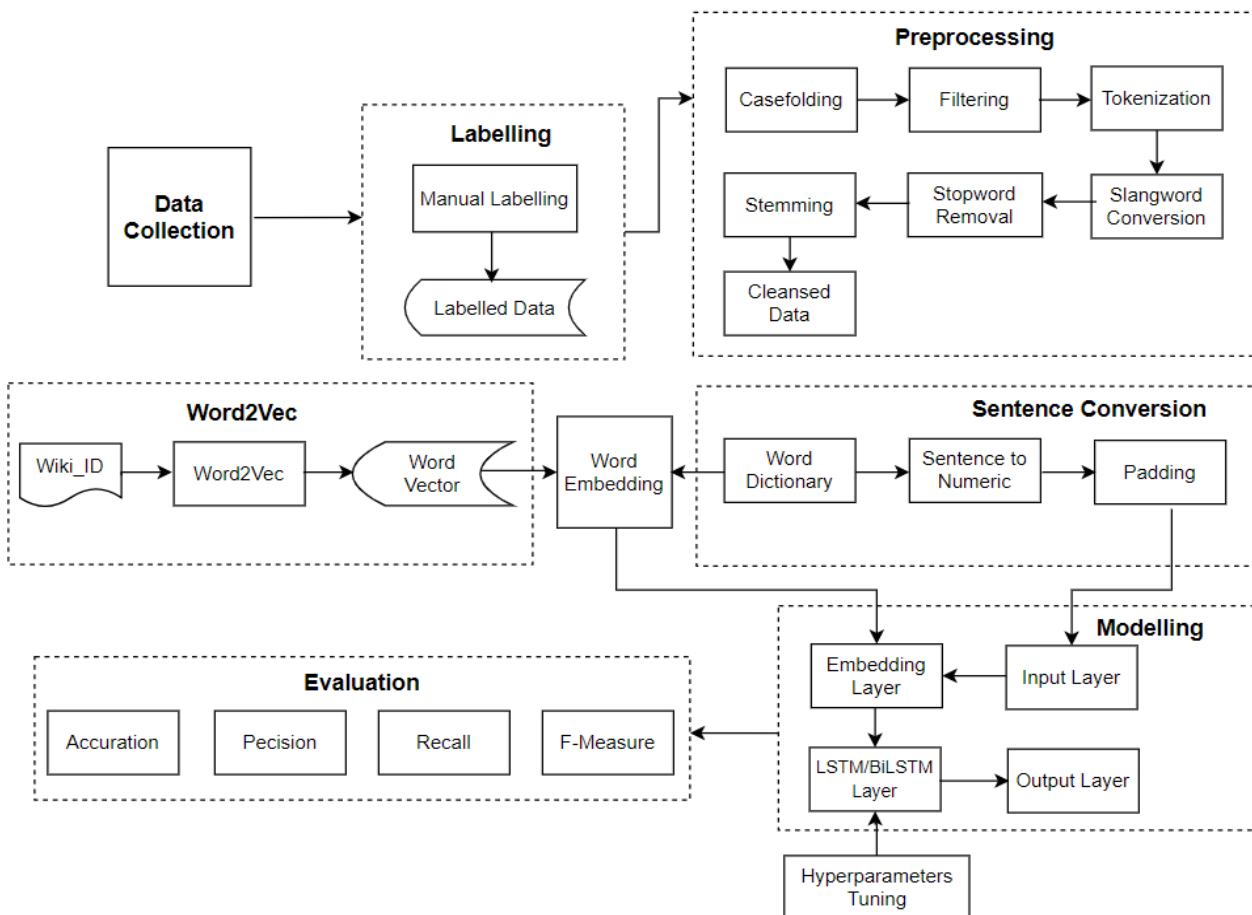


Fig. 1 Research procedure

C. Word Embedding

Word embedding is required feature extraction before the data is processed by deep learning methods (in this study: LSTM and Bi-LSTM). Word2Vec, as one of the word embedding algorithms, is chosen because it can capture the semantic meaning of the text well, and each related word is characterised by similar vectors [34]. Word embedding is divided into sentence conversion and pre-training Word2vec. Conversion of sentences starts from the word dictionary, converting sentences to numeric and padding. Meanwhile, pre-training Word2vec uses data from Indonesia Wikipedia (<http://dumps.wikimedia.org/>). The Word2vec in this study is the same as in the study [35], consisting of Continuous Bag of Word (CBOW), Hierarchical Soft-max, and 200 dimensions.

D. Hyperparameters Tuning

The hyperparameter tuning is expected to produce optimal parameters to make the best model. This process aims to find hyperparameters that provide a model of the data [36]. This Hyperparameters tuning is accomplished using the GridSearchCV library. This research's parameters include learning rate, dropout probability, and batch size. The GridSearchCV method finds a classifier's ideal parameters so that a model can correctly predict certain unlabeled data. The GridSearchCV approach is among the most effective for finding the ideal mix of the various hyperparameters in the classification model [37].

E. Modelling

This study implemented a binary relevance strategy to handle the problem of multi-label classification [38]. Therefore, the ABSA process is divided into a model generation of aspect classification and a model generation of sentiment classification. In our study, four models of aspect classification are built according to the number of aspects, and four models of sentiment classification on each aspect. Every model was constructed through LSTM and Bi-LSTM, resulting in 16 models. Each classification model is built through two approaches, namely (i) LSTM and (ii) BiLSTM.

LSTM is a Recurrent Neural Network (RNN) variant. LSTM was developed to solve the vanishing gradient problem commonly found in conventional RNNs [39]. LSTM uses three gates, namely the input gate, forget gate, and output gate, to control the use and update of the previous text information. Memory cells and three gates are designed to allow LSTMs to read, store and update previous information [39]. Bi-LSTM is a transformation over the LSTM. The properties of LSTM serialisation

processing information cause information to be processed sequentially, making it impossible to access the future context and synthesise context data, which impacts the prediction effect. While Bi-LSTM suggests using two LSTM networks to train simultaneously, one training sequence begins from the front and the other starts from the back. The two training sequences are connected to the same output layer, allowing each point's past and future information to be combined [23]-[24].

The Difference between LSTM and Bi-LSTM is that LSTM only processes data in one direction, while Bi-LSTM processes data back and forward. The LSTM training process uses one model, while the Bi-LSTM training process uses two. The first model learns the sequence of the input provided, and the second model learns the reverse of that sequence. Therefore, Bi-LSTM requires combining them [23].

F. Evaluation

Evaluation cover three stages of testing carried out; the first is testing by determining the best parameters on the LSTM and Bi-LSTM architecture. Then those parameters will be used for testing the performance of LSTM compared to Bi-LSTM. The third testing stage is to test the aspect and sentiment classification model sequentially.

First, tuning hyperparameters evaluation which is determining the best parameters of LSTM and Bi-LSTM requires tuning hyperparameters using the library GridSearchCV. This library aims to find the parameters which produce the most accuracy [37]. Second, testing models of aspect classification and sentiment classification. this study's training and testing data use a stratified cross-validation method [40] which uses $k=5$. Thus, the highest accuracy generated on each fold will be stored in a variable. While the evaluation value of models is obtained from the average accuracy, precision, recall and f1 score on each fold. Third, testing aspect classification model and sentiment classification model in sequential. The last test is to obtain sequentially the best performance of the classification and sentiment classification models. The aspect label is "1" (true) or is detected as having an attraction aspect; then, sentiment data will be called and followed by the sentiment classification stage using the sentiment attraction classification model previously generated. The output of this sentiment classification will generate sentiment labels "1" (positive) and "0" (negative). However, if the detected aspect is "0" (false) or not detected as having an attraction aspect, the label of that aspect will be changed to "-1". In this testing, the data is divided into 80% training data and 20% as test data. This distribution of

data is based on the scaling law discovered by Guyon [41]. After that, this sequential test obtains the value of accuracy, precision, recall and f1 score.

III. RESULT AND DISCUSSION

In this study, testing will be carried out by searching for the best parameters in the LSTM and Bi-LSTM architecture. Then those parameters will be used for performance testing of the LSTM architecture compared to Bi-LSTM architecture.

A. Data Distribution

This study used data reviews of 10 priority attractions crawled from tripadvisor.com in the dataset. The data is divided into two files and labelled according to the category of aspects. The first data is aspect category data labelled "1" (true) and "0" (false) to detect whether a document has an aspect category or not. The second data is sentiment data labelled "1" (positive) and "0" (negative) to detect sentiment polarity in each of the detected aspect categories. The distribution of the aspect category classification dataset in this study can be seen in Table I. Then, the data will be divided into training and test data. Meanwhile, the training data will be divided again into 80% of the training data: 20% of the testing data with validation data and the remaining data used for training data.

The labels on the sentiment classification dataset consist of positive (1), negative (0), and none (-1) sentiment polarities. The missing label (-1) in the sentiment classification dataset is a label that defines the absence of detectable aspects in a document. The distribution of sentiment classification datasets with positive (1) and negative (0) sentiment polarities can be seen in Table II.

B. Hyperparameters Tuning Evaluation

Hyperparameter tuning is a stage to search for parameters that produce the best accuracy, precision, recall, and f-measure of the LSTM architecture. The value of the hyperparameters used in tuning is a range of values considered optimal in each aspect. Then, the value will be tested on all aspect categories to find the optimal value. Tuning is done using the GridSearchCV library. This library aims to find the parameters that produce the best accuracy. Parameters can be seen in Table III.

Table IV describes the results of hyperparameters tuning to perform aspects classification using either LSTM or Bi-LSTM. For LSTM, the optimal

hyperparameters have a learning rate of 0.01, 0.001 and 0.0001, while for dropout is 0.7 and 0.2, while for batch size is 3, 32, and 64. The hyperparameters for each aspect classification can be seen in Table IV, where the hyperparameters of each aspect will be used in the LSTM training process for aspect classification.

The ideal hyperparameters for Bi-LSTM have learning rates of 0.001 and 0.0001, dropout rates of 0.5 and 0.2, and batch sizes of 32 and 64, respectively. Table V details the hyperparameters for each aspect and will be applied in the Bi-LSTM training phase for aspect classification to determine how each aspect should be classified.

The results of hyperparameter calibration for sentiment classification using either LSTM or Bi-LSTM are shown in Table V. The ideal hyperparameters for LSTM have learning rates of 0.001, dropout rates of 0.2 and 0.5, and batch sizes of 3, 32, and 64. Table V, where the hyperparameters of each aspect will be implemented in the LSTM training stage for sentiment classification in each aspect.

TABLE I
DISTRIBUTION OF ASPECT CLASSIFICATION DATASETS

Categories Aspects	Label	
	True (1)	False (0)
Attractions	4354	193
Accessibility	723	3824
Facility	836	3711
Accommodation	569	3978

TABLE II
DISTRIBUTION OF SENTIMENT CLASSIFICATION DATASETS

Categories Aspects	Polarity of Sentiment		
	None (-)	Negative (0)	Positive (1)
Attraction	193	210	4144
Accessibili	3824	215	508
Facility	3.711	190	646
Accommo	3.978	110	459

TABLE III
VALUE OF HYPERPARAMETERS

No.	Parameters	Value
1.	Learning rate	0.1, 0.01, 0.001, 0.0001
2.	Dropout	0.2, 0.5, 0,7
3.	Batch Size	3, 7, 15, 32, 64

TABLE IV
EVALUATION OF HYPERPARAMETERS TUNING FOR ASPECT CLASSIFICATION

Aspects	LSTM				Bi-LSTM			
	Learning Rate	Dropout	Batch size	Accuracy	Learning Rate	Dropout	Batch size	Accuracy
Attraction	0.0001	0.7	3	95.79%	0.001	0.2	64	95.82%
Accessibility	0.01	0.7	32	84.49%	0.001	0.5	32	84.95%
Facility	0.0001	0.2	3	83.00%	0.0001	0.2	32	82.87%
Accommodation	0.001	0.2	64	90.26%	0.001	0.2	64	90.76%

For Bi-LSTM, learning rates of 0.001, dropout rates of 0.5 and 0.2, and batch sizes of 3, 32 and 64, respectively, are the best hyperparameter values. The hyperparameters for each aspect that will be applied during the Bi-LSTM training phase for sentiment classification are listed in Table V.

C. Result of ABSA Modelling Using LSTM and Bi-LSTM

An optimal hyperparameter for aspect classification is used to conduct training on LSTM and Bi-LSTM for aspect classification. Likewise, in every aspect, optimal hyperparameters for sentiment classification are used to conduct training on LSTM and Bi-LSTM sentiment classification. The training process uses stratified cross-validation with a k = 5. The method of 5-cross validation divides the data into 5 data partitions. Furthermore, training will be carried out on the model, and an evaluation will be carried out on the model by displaying the average results of accuracy, precision, recall and f-measure for each fold. Because the training uses stratified cross-validation, detecting the highest k-fold is necessary. The model formed from the highest k-fold is loaded and stored in pickle form to be used in the following process. The next process is to load 4 LSTM models and 4 Bi-LSTM models for aspect classification. Also, add load 4 LSTM models and 4 Bi-LSTM models for sentiment classification. Therefore, this modelling process produces 16 models, as shown in Table II.

D. Aspect Classification Evaluation

Tuning hyperparameters is a stage to search for parameters that produce the best accuracy, precision, recall, and f-measure of the LSTM and Bi-LSTM

architecture. The value of the hyperparameters used in tuning is a range of considered values.

The summary of the aspect classification test result is depicted in Table VI. Aspect category classification testing using LSTM and Bi-LSTM yielded precision, recall, and f1 score values. The best F1 score was obtained in testing aspects of attraction in classifying aspects using Bi-LSTM. F1 score for accessibility, facility and accommodation on the classification of aspects using Bi-LSTM got lower results. The best value of the F1 score is 97.99% on attraction aspects using Bi-LSTM. The test results showed that Bi-LSTM performs better than LSTM in aspect classification. Bi-LSTM obtained average accuracy values, and the f1 score outperformed 89,49% and 63,66%. While average values of accuracy and f1 score using LSTM are 89,23% and 60,49%.

E. Sentiment Classification Evaluation

The summary of the sentiment classification test result is depicted in Table VII. Sentiment classification testing using LSTM and Bi-LSTM yielded precision, recall, and f1 score values. The best deal of F1 score was obtained in testing aspects of accessibility in the classification of the sentiment using Bi-LSTM. Aspects of attractions, facilities, and accommodation on the classification of aspects using Bi-LSTM got lower results. The best value of the F1 score is 61.86% on accessibility aspects using Bi-LSTM. The best micro average on sentiment classification is 55.78% using Bi-LSTM. Table VII shows that Bi-LSTM can improve the F1 score by 1.34% and the accuracy value by 2%.

TABLE V
EVALUATION OF HYPERPARAMETERS TUNING FOR SENTIMENT CLASSIFICATION

Aspects	LSTM				Bi-LSTM			
	Learning Rate	Dropout	Batch size	Accuracy	Learning Rate	Dropout	Batch size	Accuracy
Attraction	0.001	0.5	3	95.20%	0.001	0.2	3	95.00%
Accessibility	0.001	0.2	32	70.58%	0.001	0.5	32	70.76%
Facility	0.001	0.5	3	78.59%	0.001	0.5	32	77.39%
Accommodation	0.001	0.2	64	83.08%	0.001	0.2	3	81.54%

TABLE VI
EVALUATION OF ASPECT CLASSIFICATION

Aspects	Aspect Category Classification (LSTM)				Aspect Category Classification (Bi-LSTM)			
	Test Accuracy	Precision	Recall	F-1	Test Accuracy	Precision	Recall	F-1
Attraction	95.93%	96.56%	99.31%	97.92%	96.09%	96.47%	99.57%	97.99%
Accessibility	83.85%	49.15%	19.86%	28.29%	86.06%	64.02%	37.57%	46.41%
Facility	85.37%	64.60%	43.98%	52.33%	84.90%	63.76%	43.53%	51.61%
Accommodation	91.76%	68.42%	59.09%	63.41%	90.90%	67.10%	52.12%	58.62%
Micro Average	89.23%	69.68%	55.56%	60.49%	89.49%	72.84%	58.20%	63.66%

TABLE VII
EVALUATION OF ASPECT CLASSIFICATION

Aspects	Sentiment Classification (LSTM)				Sentiment Classification (Bi-LSTM)			
	Test Accuracy	Precision	Recall	F-1	Test Accuracy	Precision	Recall	F-1
Attractions	94.60%	94.60%	94.60%	51.96%	96.10%	96.10%	86.44%	57.03%
Accessibility	67.59%	67.59%	67.59%	56.33%	73.79%	73.79%	69.48%	61.86%
Facility	76.19%	76.19%	76.19%	52.05%	75.60%	75.60%	66.48%	51.80%
Accommodation	79.82%	79.82%	89.55%	57.41%	80.70%	80.70%	68.52%	52.44%
Micro Average	79.55%	79.55%	81.98%	54.44%	81.55%	81.55%	72.73%	55.78%

F. Evaluation of Aspect Classification and Sentiment Classification Sequentially

Sequential analysis of the classification of aspect and sentiment categories is a test carried out in a multilevel manner with sequential analysis load against aspect and sentiment models trained in the previous stage. The classification of aspect categories is carried out first and then continued with the type of sentiments.

The test results on the classification of aspect categories and sentiment sequentially in some aspects also produce precision, recall, and f-measure, as shown in Table VIII. The test results showed that Bi-LSTM performs better than LSTM in aspect and sentiment classification sequentially. Bi-LSTM attained average accuracy values, and the f1 score outperformed 92,22% and 71,06%. While average values of accuracy and f1 score using LSTM are 90,63% and 70,42%.

G. Comparison of LSTM Results and Bi-LSTM Results

Based on the research, aspect-based sentiment analysis using the Bi-LSTM method performed quite well in aspect classification. The calculation of classification performance using micro averages obtained accuracy, and the f1 score results in the aspect classification using Bi-LSTM were 89.49% and 63.66%, respectively, better than LSTM. Meanwhile, in the sentiment classification using LSTM, the accuracy and f1 score value was slightly lower than that of Bi-LSTM. Then, aspect and sentiment classification sequentially evaluate that Bi-LSTM obtained better accuracy and f1 score than LSTM. This shows that Bi-LSTM is more powerful than LSTM in ABSA research on Indonesian tourist attraction reviews. Attraction reviews on tripadvisor.com are usually in the form of long sentences, thus requiring a Bi-LSTM that can recall data back and forth.

TABLE VIII
EVALUATION OF ASPECT CLASSIFICATION AND SENTIMENT CLASSIFICATION SEQUENTIALLY

Aspects	Sequential LSTM				Bi-LSTM			
	Test Accuracy	Precision	Recall	F-1	Test Accuracy	Precision	Recall	F-1
Attractions	93.74%	93.98%	99.52%	96.67%	94.50%	94.55%	99.76%	97.08%
Accessibility	86.15%	64.29%	31.58%	42.35%	89.34%	56.86%	37.17%	44.96%
Facility	86.26%	55.98%	59.53%	57.70%	91.75%	73.43%	75.20%	74.30%
Accommodation	96.37%	86.32%	84.97%	84.97%	93.29%	74.32%	62.50%	67.90%
Micro Average	90,63%	75,14%	68,58%	70,42%	92,22%	74,79%	68,66%	71,06%

IV. CONCLUSION

Our research proposed Bi-LSTM to conduct an ABSA task based on reviews of 10 prioritised tourist attractions with a binary relevance mechanism. On aspect and simultaneous sentiment classification, Bi-LSTM obtained an accuracy value of 92.22% and an f1 score of 71.06%. This research also produces the eight best Bi-LSTM aspect and sentiment, classification models. The parameters of this study's eight best Bi-LSTM models have also been optimised using the hyperparameters tuning method. Furthermore, these models can be implemented to perform ABSA tasks on other tourist attraction review data. Meanwhile, another concern is that all models obtain high accuracy above 70% but still have lower F1 scores, hence using an imbalanced dataset. Future research should be conducted on ABSA for this dataset by using the method of handling imbalanced data.

ACKNOWLEDGEMENT

The authors thank The Directorate of Research, Technology and Higher Education, the Republic of Indonesia, through the PDP program Fund 2022.

REFERENCES

- [1] A. Haryana, "Economic and Welfare Impacts of Indonesia's Tourism Sector," *J. Perenc. Pembang. Indones. J. Dev. Plan.*, vol. 4, no. 3, pp. 300–311, 2020, doi: 10.36574/jpp.v4i3.127.
- [2] D. M. Lemy, F. Teguh, and A. Pramezwarly, "Tourism Development in Indonesia," vol. 11, pp. 91–108, 2019, doi: 10.1108/s2042-144320190000011009.
- [3] M. . Dr. M. Agus Cholik, S.E, "the Development of Tourism Industry in Indonesia.," *Eur. J. Res. Reflect. Manag. Sci.*, vol. 5, no. 1, pp. 49–59, 2017, [Online]. Available: www.idpublications.org
- [4] P. Del Vecchio, G. Mele, V. Ndou, and G. Secundo, "Creating value from Social Big Data: Implications for Smart Tourism Destinations," *Inf. Process. Manag.*, vol. 54, no. 5, pp. 847–860, 2018, doi: 10.1016/j.ipm.2017.10.006.
- [5] B. Liu, "Sentiment analysis: Mining opinions, sentiments, and emotions," *Sentim. Anal. Min. Opin. Sentim. Emot.*, no. May, pp. 1–367, 2015, doi: 10.1017/CBO9781139084789.
- [6] S. Wu, Y. Xu, F. Wu, Z. Yuan, Y. Huang, and X. Li, "Aspect-based sentiment analysis via fusing multiple sources of textual knowledge ☆," vol. 183, p. 104868, 2019, doi: 10.1016/j.knosys.
- [7] X. Wang, X. Chen, M. Tang, T. Yang, and Z. Wang, "Aspect-Level Sentiment Analysis Based on Position Features Using Multilevel Interactive Bidirectional GRU and Attention Mechanism," *Discret. Dyn. Nat. Soc.*, vol. 2020, 2020, doi: 10.1155/2020/5824873.
- [8] S. Gojali and M. L. Khodra, "Aspect based sentiment analysis for review rating prediction," in *2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA)*, 2016, pp. 1–6. doi: 10.1109/ICAICTA.2016.7803110.
- [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)*, 2017, pp. 1–6. doi: 10.1109/ICAICTA.2017.8090963.
- [10] 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. Cybern.*, vol. 10, no. 8, pp. 2163–2175, Aug. 2019, doi: 10.1007/s13042-018-0799-4.
- [11] H. Gandhi and V. Attar, "Extracting Aspect Terms using CRF and Bi-LSTM Models," in *Procedia Computer Science*, 2020, vol. 167, pp. 2486–2495. doi: 10.1016/j.procs.2020.03.301.
- [12] A. S. Shafie, N. M. Sharef, M. A. A. Murad, and A. Azman, "Aspect Extraction Performance with POS Tag Pattern of Dependency Relation in Aspect-based Sentiment Analysis," in *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, 2018, pp. 1–6. doi: 10.1109/INFRKM.2018.8464692.
- [13] Z. Fachrina and D. Widyantoro, *Aspect-sentiment classification in opinion mining using the combination of rule-based and machine learning*. 2017. doi: 10.1109/ICODSE.2017.8285850.
- [14] S. Cahyaningtyas, D. Hatta Fudholi, and A. Fathan Hidayatullah, "Deep Learning for Aspect-Based Sentiment Analysis on Indonesian Hotels Reviews," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, Aug. 2021, doi: 10.22219/kinetik.v6i3.1300.
- [15] N. Jihan, Y. Senarath, D. Tennekoon, M. Wickramarathne, and S. Ranathunga, "Multi-Domain Aspect Extraction using Support Vector Machines Tamil News Clustering Using Word Embeddings View project Aspect Based Sentiment Analysis of Customer Reviews View project Multi-Domain Aspect Extraction using Support Vector Machines." [Online]. Available: <https://www.researchgate.net/publication/322314286>
- [16] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews," *J. Comput. Sci.*, vol. 27, pp. 386–393, Jul. 2018, doi: 10.1016/j.jocs.2017.11.006.

- [17] L. P. Manik *et al.*, “Aspect-Based Sentiment Analysis on Candidate Character Traits in Indonesian Presidential Election,” pp. 224–228, 2020.
- [18] S. M. E and R. Sunitha, “Survey On Aspect Based Sentiment Analysis Using Machine Learning Techniques,” vol. 07, no. 10, 2020.
- [19] K. Shuang, X. Ren, Q. Yang, R. Li, and J. Loo, “AELA-DLSTMs: Attention-Enabled and Location-Aware Double LSTMs for Aspect-level Sentiment Classification,” *Neurocomputing*, 2018, doi: 10.1016/j.neucom.2018.11.084.
- [20] K. Srividya and A. M. Sowjanya, “Materials Today : Proceedings NA-DLSTM – A neural attention based model for context aware Aspect-based sentiment analysis,” *Mater. Today Proc.*, no. xxxx, 2021, doi: 10.1016/j.matpr.2021.01.782.
- [21] A. Yadav and D. K. Vishwakarma, “Sentiment analysis using deep learning architectures: a review,” *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4335–4385, 2020, doi: 10.1007/s10462-019-09794-5.
- [22] A. Aziz Sharfuddin, M. Nafis Tihami, and M. Saiful Islam, “A Deep Recurrent Neural Network with BiLSTM model for Sentiment Classification,” *2018 Int. Conf. Bangla Speech Lang. Process. ICBSLP 2018*, pp. 1–4, 2018, doi: 10.1109/ICBSLP.2018.8554396.
- [23] K. Zhang, W. Song, L. Liu, X. Zhao, and C. Du, “Bidirectional long short-term memory for sentiment analysis of Chinese product reviews,” *ICEIEC 2019 - Proc. 2019 IEEE 9th Int. Conf. Electron. Inf. Emerg. Commun.*, pp. 665–668, 2019, doi: 10.1109/ICEIEC.2019.8784560.
- [24] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, “Sentiment analysis of comment texts based on BiLSTM,” *IEEE Access*, vol. 7, no. c, pp. 51522–51532, 2019, doi: 10.1109/ACCESS.2019.2909919.
- [25] A. Suciati and I. Budi, “Aspect-Based Sentiment Analysis and Emotion Detection for Code-Mixed Review,” vol. 11, no. 9, pp. 179–186, 2020.
- [26] A. Cahyadi and M. L. Khodra, “Aspect-Based Sentiment Analysis Using Convolutional Neural Network and Bidirectional Long Short-Term Memory,” *2018 5th Int. Conf. Adv. Informatics Concept Theory Appl.*, pp. 124–129, 2018.
- [27] M. Theo, A. Bangsa, S. Priyanta, and Y. Suyanto, “Aspect-Based Sentiment Analysis of Online Marketplace Reviews Using Convolutional Neural Network,” vol. 14, no. 2, pp. 123–134, 2020, doi: 10.22146/ijccs.51646.
- [28] A. A. Fattahila, F. I. Amorokhman, K. M. Auditama, K. A. Wijaya, and A. Romadhony, “Indonesian Digital Wallet Sentiment Analysis Using CNN And LSTM Method,” in *2021 International Conference on Artificial Intelligence and Big Data Analytics*, 2021, pp. 1–6. doi: 10.1109/ICAIBDA53487.2021.9689712.
- [29] M. R. Yanuar and S. Shiramatsu, “Aspect Extraction for Tourist Spot Review in Indonesian Language using BERT,” in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIC)*, 2020, pp. 298–302. doi: 10.1109/ICAIC48513.2020.9065263.
- [30] D. Arianto, “Aspect- based Sentiment Analysis on Indonesia ’ s Tourism Destinations Based on Google Maps User Code-Mixed Reviews (Study Case : Borobudur and Prambanan Temples),” vol. 2019, 2019.
- [31] L. C. Dewi, Meiliana, and A. Chandra, “Social media web scraping using social media developers API and regex,” *Procedia Comput. Sci.*, vol. 157, pp. 444–449, 2019, doi: 10.1016/j.procs.2019.08.237.
- [32] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, “A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis,” *Expert Syst. Appl.*, vol. 110, pp. 298–310, 2018, doi: 10.1016/j.eswa.2018.06.022.
- [33] J. Asian, H. E. Williams, and S. M. M. Tahaghoghi, “Stemming Indonesian,” *Conf. Res. Pract. Inf. Technol. Ser.*, vol. 38, no. 4, pp. 307–314, 2005, doi: 10.1145/1316457.1316459.
- [34] Md Al-Amin, M. S. Islam, and S. Das Uzzal, “Sentiment analysis of Bengali comments with Word2Vec and sentiment information of words,” *ECCE 2017 - Int. Conf. Electr. Comput. Commun. Eng.*, pp. 186–190, 2017, doi: 10.1109/ECACE.2017.7912903.
- [35] D. I. Afidah, R. Kusumaningrum, and B. Surarso, “Long short term memory convolutional neural network for Indonesian sentiment analysis towards touristic destination reviews,” *Proc. - 2020 Int. Semin. Appl. Technol. Inf. Commun. IT Challenges Sustain. Scalability, Secur. Age Digit. Disruption, iSemantic 2020*, pp. 630–637, 2020, doi: 10.1109/iSemantic50169.2020.9234210.
- [36] B. H. Shekar and G. Dagnev, “Grid search-based hyperparameter tuning and classification of microarray cancer data,” *2019 2nd Int. Conf. Adv. Comput. Commun. Paradig. ICACCP 2019*, pp. 1–8, 2019, doi: 10.1109/ICACCP.2019.8882943.
- [37] C. G. Siji George and B. Sumathi, “Grid search tuning of hyperparameters in random forest classifier for customer feedback sentiment prediction,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 9, pp. 173–178, 2020, doi: 10.14569/IJACSA.2020.0110920.
- [38] G. Wu, Y. Tian, and C. Zhang, “A unified framework implementing linear binary relevance for multi-label learning,” *Neurocomputing*, vol. 289, pp. 86–100, 2018, doi: 10.1016/j.neucom.2018.02.010.
- [39] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [40] G. Jiang and W. Wang, “Error estimation based on variance analysis of k-fold cross-validation,” *Pattern*

- Recognit.*, vol. 69, pp. 94–106, 2017, doi: 10.1016/j.patcog.2017.03.025.
- [41] I. Guyon, “A scaling law for the validation-set training-set size ratio,” *AT&T Bell Lab.*, pp. 1–11, 1997, [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.33.1337&rep=rep1&type=pdf>