



Deep learning for aspect-based sentiment analysis on Indonesian hotels reviews

Siwi Cahyaningtyas^{1*}, Dthomas Hatta Fudholi², Ahmad Fathan Hidayatullah³

Master Program in Informatics, Department of Informatics, Universitas Islam Indonesia, Indonesia¹

Department of Informatics, Universitas Islam Indonesia, Indonesia²

Department of Informatics, Universitas Islam Indonesia, Indonesia³

Article Info

Keywords:

Aspect-based Sentiment Analysis, Hotel, Deep Learning

Article history:

Received: July 11, 2021

Accepted: August 05, 2021

Published: August 31, 2021

Cite:

Cahyaningtyas, S., Hatta Fudholi, D., & Fathan Hidayatullah, A. (2021). Deep Learning for Aspect-Based Sentiment Analysis on Indonesian Hotels Reviews. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 6(3).

<https://doi.org/10.22219/kinetik.v6i3.1300>

*Corresponding author

Siwi Cahyaningtyas

E-mail address:

19917035@students.uii.ac.id

Abstract

Tourism is one of the fastest-growing industries. Many travelers book hotels and share their experiences using travel e-commerce sites. To improve the quality of products and services, we can take advantage by analyzing their reviews. We can see the good and the bad thing reviews in every aspect of the hotel. However, research to analyze sentiment in every aspect using Indonesian hotel reviews is still relatively new. In this work, we propose to create an Aspect-based Sentiment Analysis (ABSA) using Indonesian hotel reviews to solve the problem. This research consists of four steps: collecting data, preprocessing, aspect classification, and sentiment classification. Our classification process compares with eight deep learning methods (RNN, LSTM, GRU, BiLSTM, Attention BiLSTM, CNN, CNN-LSTM, and CNN-BiLSTM). In aspect classification, we have six classes of aspects which are *harga* (price), *hotel*, *kamar* (room), *lokasi* (location), *pelayanan* (service), and *restoran* (restaurant). In sentiment analysis, we compared two scenarios to classify sentiments as positive or negative. The first one is to classify sentiment in all aspects, and the second one is to classify sentiment in every aspect. The results showed that LSTM achieved the best model for aspect classification with an accuracy value of 0.926. For sentiment classification, our experiments showed that classify sentiment in every aspect achieved a better result than classify sentiment in all aspects. The result showed that the CNN model gets an average accuracy score of 0.904.

1. Introduction

Tourism is an industry that is the key to a country's economic growth. The travelers book hotels online and share their experiences in an online platform or so-called User-generated content (UGC). UGC is an online platform for giving ratings and reviews of a product from customers [1]. The reviews are considering to be able to provide indirect experience to others customers. A survey by Ady and Quadri [2] states that 93% of people researching by looking at all reviews before choosing a product. Besides the customer side, reviews on UGC also give benefits on the marketing business side. Reviews with positive comments consider helping promotion indirectly, especially when knowing the attribute in every sentiment [3]. To get the insights of sentiment in every review, this study proposes to use aspect-level sentiment analysis.

Aspect-level sentiment analysis or so-called Aspect-based Sentiment Analysis (ABSA) is a type of sentiment analysis that can see all of the sentiment in every aspect. ABSA is done in various ways. In research by [4][5][6], they are using aspect extraction process with POS Tagger to extract aspect and opinion word before to classify in aspect and sentiment. But in some studies [7][8][9], aspect extraction is not used for the ABSA process because it requires a lot of resources [10]. Other than that, step on sentiment classification in ABSA also can be done in various ways. We can solve the sentiment classification in ABSA by classifying all aspects [11] or every aspect [8].

ABSA process applied in several domains and languages. As in product reviews in Chinese [7] that using comparison of SVM, LSTM, Bi-LSTM, TD-LSTM, TC-LSTM, and ATSM-S with the best result on the ATSM-S model. Restaurant reviews in English [12] predicting 294,034 with SVM as the best model with an f1-score value of 0.7959. In hotel reviews in Arabic, [11] applied the CNN model on 4,802 reviews with an accuracy value of 0.8270.

Research on ABSA using hotel data is used in several studies with various languages, aspects, and methods. With the different characteristics of sentences in every language, the ABSA modeling process needs to adjust to the structures of the language used. In a country that has several popular destination places, ABSA in Indonesia has become an important task to develop. However, ABSA in hotel reviews in Indonesian is still relatively new. Setiowati et al. [13] predicting 321 data reviews using a comparison of Naïve Bayes, KNN, J48, SVM, and Rule-based methods. They use nine labels for aspects: engineering, f & b (food and beverages), security, housekeeping, family vacations, hotel locations, services, sales, and unknown. For sentiment classification, they have two classes which are positive

and negative. The results of aspect classification in Setiowati et al. [14] show that Rule-based get the best result with a value of 0.98. In sentiment classification, they get the Rule-based method as the highest score with a value of 0.99. Despite gets a good result, classification tasks using the Rule-based consider a lot of manual work, time-consuming, less learning capacity, and complex domains [15]. Prameswari et al. [16] proposed Recursive Neural Tensor Network (RNTN) method to classify 104 hotel reviews in Bali. They use eight classes for aspect classification that are accessibility, activities & entertainment, f&b operations, guest's perspective, HR, room amenities, transportation services, and physical environment. The result in Prameswari et al. [16] shows that the RNTN model gets the best model with an average accuracy sentiment of 0.85. The positive and negative class has an accuracy value of 0.90 and 0.64 respectively. The low negative class accuracy is due to the amount and imbalance of the training data.

According to described problems, this study aims to create an Aspect-based Sentiment Analysis (ABSA) in Indonesian hotel reviews. This research consists of four steps: collecting data, preprocessing, aspect classification, and sentiment analysis. In addition, we also compared two scenarios in sentiment analysis. The first one is to classify sentiment in all aspects, and the second one is to classify sentiment in every aspect. In this research, we strengthen our contribution by validating the dataset labels by an expert who has worked in the hospitality sector for two years, using a larger and balance dataset than previous studies to improve the accuracy, and classifies data with more general aspects (price, hotel, room, location, service, and restaurant). We adopt the price, room, location, service, and restaurant from the previous work [13][16]. In addition, we also use the hotel aspect that is the combination of aspect guest's perspective and physical environment in Prameswari et al. [16]. Besides that, the hotel aspect is also one of the essential hotel attributes based on the last two decades [17].

In the classification process, this study uses a comparison of deep learning methods (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). Although the Rule-based model gets almost perfect accuracy [15], that method requires a lot of resources. In addition, deep learning is also proving to deal with the classification process with a large amount of data [7][18].

2. Research Method

Aspect-based Sentiment Analysis (ABSA) is a type of sentiment analysis at the aspect level. In this work, we adapt from the previous study by [10] without using the POS Tagging process. Our ABSA process consists of four steps that show in Figure 1.



Figure 1. Research Step

The first step is to collect data using Selenium Webdriver packages in Python. Collect data in our study use scrap technique from hotel reviews like Agoda, Booking.com, PegiPegi, Traveloka, and Tiket.com which are the 15 most popular travel apps in Indonesia in Google PlayStore [19]. The data is then manually labeled by following the previous study by [17] and verified by an expert who has worked in the hospitality sector for two years. The attributes of the hotel aspect that used in this study show in Table 1.

Table 1. Attributes of the Hotel Aspect

Aspect	Item
Price	Price of accommodation, value for money, discount rate, free parking, free newspaper, free breakfast, f & b prices, free local calls, free cable tv
Hotel	Architecture, hotel decoration, cleanliness, aesthetics, elevators, legible, visible signs in public areas, wide doorways, carpet, public space, bar, shops, games room, parking facilities, fitness facilities, swimming pool, sauna, meeting rooms, fax machine, lighting equipment
Room	Standard of bedrooms, cleanliness, size, comfort, aesthetics, quiet/soundproof rooms, well-maintained furnishings, non-smoking, comfort of the bed, bath size, bathroom amenities, hot water supply, bathroom furniture, bathroom cleanliness, good quality towels, tv/radio.
Location	Convenient, convenient to parking, convenient to the airport, convenient to downtown, convenient to business, well-it public areas, quiet area
Service	Check-in and check-out speed, friendliness of staff, service professionalism, service speed, service efficiency, laundry service, valet parking, 24-hour room service, reservation handling, complaint responsiveness.
Restaurant	Restaurant facilities, gourmet/specialty restaurants, 24-hour coffee shop, quality, atmosphere, food service quality, variety, dietary menus, bar, hygiene, food portions.

After labeling data, the second step is the preprocessing data. The preprocessing tasks in this study adopt from our previous study [18] are removed punctuations, removed non-ASCII characters, removed URLs, removed digits, removed whitespaces, lowercase, normalization, stopword, and join negation. Join negation use to connect the negation words with an underscore punctuation mark (“_”). For the filtering process, we use “re” and “nltk” libraries. Besides that, we add our data for the normalization and stopword process.

The next step is aspect classification. In aspect classification, we have six classes: price, hotel, location, price, restaurant, room, and service. In sentiment analysis, we compared two scenarios to classify sentiments as positive or negative. The first scenario is to classify sentiment in all aspects, and the second one is to classify sentiment in every aspect. In the first scenario, we use the highest accuracy score for the evaluation results. While in the second scenario, we calculate the average score from all aspects in each model for the evaluation results [8]. The highest score in the two experiments compares to get the best result.

These classifications processes divide the dataset into the training data (2/3) and the evaluation data (1/3). Our classification process compares deep learning methods and hyperparameter values by following the previous study [20]. In addition, we use categorical cross-entropy for loss function and softmax for activation to handle aspect classification. For sentiment classification, we use binary categorical cross-entropy for loss function and softmax for activation. For the number of epochs, we use the early stopping function. The early stopping function defines the number of epochs when the performance model has stopped improving. For the method, we use RNN, LSTM, GRU, BiLSTM, Attention BiLSTM, CNN, CNN-LSTM, and CNN-BiLSTM that applied in our previous study [18]. To create the model, we use Keras packages from Python.

2.1 Recurrent Neural Network (RNN)

RNN is a type of neural network designed to recognize sequential data for ordinal problems [21]. RNN is suitable for sequential data such as language translation, speech recognition, and image captioning because they use the same function for every input. Unlike the other neural network, RNN has internal memory to remember the important things about the previous inputs [22].

Figure 2 shows the “rolled” visual of the RNN represents all the information, and the “unrolled” represents the individual layers of the neural network. The unrolled version indicates that multiple nodes are connected by passing information from one state to another. The first input (X_0) computes the function on hidden layers and gives the output (Y_0). Output from the current state (Y_0) is an input in the next state (X_1) to produce an output from that state (Y_1).

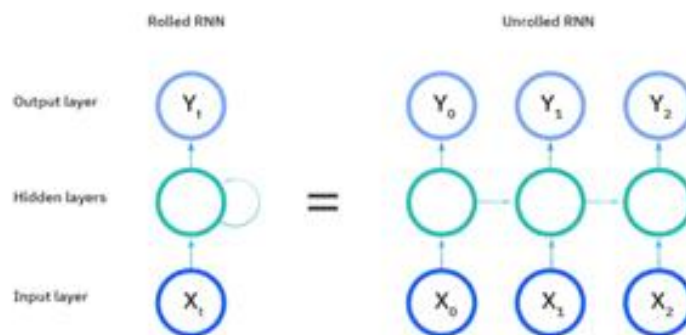


Figure 2. Architecture of RNN

In our training model, first, we use the embedding layer to convert text into vectors. In the next layer, we use the dropout layer to reduce the overfitting. Next step we use the RNN layer by importing Keras layers. In the last layer, we use dense to activate the neuron use activation function.

2.2 Long Short-term Memory (LSTM)

LSTM is a modified version of RNN that can handle the vanishing gradient problem using backpropagation. The vanishing gradient problem in RNN causes its inability to learning data with long-term dependencies. A single LSTM cell can be computed as follows Equation 1, Equation 2, Equation 3, Equation 4, and Equation 5 [23].

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (3)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{4}$$

$$h_t = o_t \tanh(c_t) \tag{5}$$

An LSTM model consists of a cell state and three gates to decide how much the information will be kept and throw away. The input gate (i_t) updates the cell state bypassing the current input (x_t) and previously hidden state (h_{t-1}) into the sigmoid function (σ) and transform it to the range 0 – 1. The f_t is a forget gate that determines the amount of memory that needs to throw away, and the o_t is an output gate that decides what the next hidden state should be. The cell state represents with c_t to store the information from the new state by computing the weighted sum from the previous cell state and current information generated by the cell. The h_t is a hidden layer state at time t by computing o_t over this updated history c_t using tangent function [11]. The LSTM in this work consists of embedding, dropout, and dense layers.

2.3 Gated Recurrent Unit (GRU)

GRU is a modified version of LSTM that can capture long-term dependencies without removing irrelevant information with simpler architecture. GRU only uses a cell state and two kinds of gates which are update and reset gates. The update gate (z_t) helps determine how much information to keep, and the reset gate (r_t) decides how much the past information to forget. The candidate state (\hat{h}_t) will use the reset gate to store information from the past, and then the result will be represented as an output vector using a hidden state (h_t) [24]. In our study, we use embedding, dropout, GRU, and dense layers. The basic formula of GRU are as follows Equation 6, Equation 7, Equation 8, and Equation 9 [24].

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{6}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{7}$$

$$\hat{h}_t = \tanh(W_h x_t + U_h (r_t h_{t-1})) + b_h \tag{8}$$

$$h_t = (1 - z_t)h_{t-1} + z_t \hat{h}_t \tag{9}$$

2.4 Bidirectional LSTM (BiLSTM)

BiLSTM is an improvement from the LSTM model that can handle the contextual information from the future token [25]. The basic idea of BiLSTM is to obtain the contextual information by combines two hidden layers, forward (\vec{h}_t) and backward (\overleftarrow{h}_t) hidden layers [26] from opposite directions to the same output (y_t) that shows in Figure 3. By using the forward and backward layers, BiLSTM can capture the past and future information. Architecture BiLSTM in our study consists of four layers: embedding, dropout, BiLSTM, and dense layers.

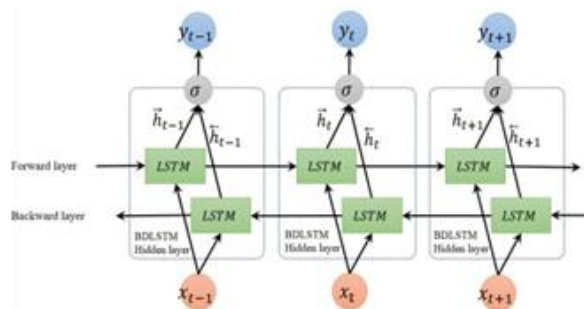


Figure 3. Architecture of BiLSTM

2.5 Attention BiLSTM

Attention BiLSTM is a combination of the Attention layer and BiLSTM model. Attention layer is a type of neural network that has recent success in image and text data. Attention layer proves to solve in text classification like sentiment classification [27][28]. Sentiment classification captures the correlation between words and calculates the important words by the weight in every word. The higher weight, the more important the word is [23]. The architecture of Attention BiLSTM in this work consists of the input layer, embedding layer, BiLSTM layer, attention layer, and an output layer that apply in the previous study [29].

2.6 Convolutional Neural Network (CNN)

CNN is a deep learning algorithm that commonly uses in computer vision and image processing like image classification, object detection, and others [21]. Currently, more studies are actively using CNNs in text classification. The most popular CNN model for text classification is done by [30]. CNN model commonly consists of convolutional, pooling, and fully connected layers. To deal with the text classification problems, our CNN model adds the embedding layer to convert the input text into a fixed-size vector [21]. In the next layer, we use convolutional and pooling layers to handle feature extraction and the fully connected layers map to extract the features into the final output in the last layer.

2.7 CNN-LSTM

CNN LSTM is a combination of CNN layers to extract the feature on input data and LSTM layers to support sequence prediction. CNN LSTM commonly uses to solved sequence prediction problems with spatial input like images or videos [31]. In addition, the CNN LSTM has also proved to solve the sequence data in the previous study that classify text for sentiment and emotion [32]. The architecture of CNN-LSTM consists of combining CNN and LSTM layers. In this work, we use embedding, convolution, pooling, fully connected layer, dense, and the LSTM layer.

2.8 CNN-BiLSTM

CNN-BiLSTM is a type of neural network that combining CNN and BiLSTM models. This model applies by combining CNN layers for feature extraction and BiLSTM for the output of the max-pooling layer used to consider long-term dependency. Architecture CNN-BiLSTM has proved good results in natural language processing for sentiment analysis [33][34]. In our training model for CNN-BiLSTM, we use embedding, convolution, pooling, fully connected layer, and dense layers that use in the CNN model. To combine the CNN and BiLSTM model, we add the BiLSTM layer before the last layer.

3. Results and Discussion

This section will describe the research steps that applied in this study. In this work, we conduct it in four steps. The first step is collecting data, the second step is preprocessing, the third is aspect classification, and the last step is sentiment classification.

3.1 Data Collection

In this process, we collected 5,387 reviews by scrapping several sites such as Agoda, Booking.com, PegiPegi, Traveloka, and Tiket.com. The data is labeling with price, hotel, room, location, service, and restaurant in aspect, and positive and negative for sentiment classification for the training data. The example of training data shows in Table 2.

Table 2. Examples of Training Data

Text	Aspect	Sentiment
<i>Acnya ga dingin trs airnya juga netes trus ke lantai</i> (The AC isn't cool n some water keep continue fall to the floor)	Room	Negative
<i>Resepsionisnya baik banget, helpful, proses cekin cekout lumayan cepett</i> (The receptionist was very nice, helpful, and the check-in check-out process is quite fast)	Service	Positive
<i>Lokasi dekat banget sama malioboro tinggal jalan kaki</i> (The location is near to Malioboro so we just walked there)	Location	Positive

From the data, the result of the distribution of data labeling shows in Table 3. The distribution results show that the aspect and sentiment data have fairly balanced data. The aspect distribution has a data range of 861-963 with the lowest distribution owned by the price aspect. Meanwhile, the distribution of sentiment data consists of 2,739 and 2,648 positive and negative data.

Table 3. Label Distribution of Data that were Annotated

Aspect	Positive	Negative	Total
Price	440	421	861
Hotel	494	442	936
Room	434	529	963
Location	495	413	918
Service	438	423	861
Restaurant	438	420	858
Total	2739	2648	5387

3.2 Data Preprocessing

The review data in Table 2 shows that the data has many informal languages. This problem can deal with the preprocessing by filter the terms of unnecessary things and normalize them to a more uniform sequence. The result of preprocessing tasks shows in Table 4.

Table 4. Result of Preprocessing Task

Original Text	Preprocessing
<i>Acnya ga dingin trs airnya juga netes trus ke lantai</i> (The AC isn't cool n some water keep continue fall to the floor)	<i>ac tidak_dingin air menetes lantai</i> (ac not_cool waterfall down floor)
<i>Resepsionisnya baik banget, helpful, proses cekin cekout lumayan cepett</i> (The receptionist was very nice, helpful, and the check-in check-out process is quite fastt)	<i>resepsionis baik helpful proses checkin checkout cepat</i> (the receptionist nice helpful checkin checkout process fast)
<i>Lokasi dekat banget sama malioboro tinggal jalan kaki</i> (The location is near to Malioboro so we just walked there)	<i>lokasi dekat malioboro hanya jalan kaki</i> (location near Malioboro just walked there)

3.3 Aspect Classification

In aspect classification, we use some experiments on the architectures and hyperparameters. Each model evaluates to get the best result that shows in Table 5. After training some models, we get the accuracy values in each model are not too much different. The Attention BiLSTM has the lowest score with an accuracy value of 0.896, and the LSTM model has the highest score with an accuracy value of 0.926.

Table 5. Result of Aspect Classification

Model	Number of Units	Batch Size	Number of Filters	Kernel Size	Dropout	Accuracy
RNN	300	32	-	-	0.3	0.901
LSTM	100	32	-	-	0.5	0.926
GRU	100	256	-	-	0.3	0.922
BiLSTM	100	128	-	-	0.5	0.902
Attention BiLSTM	100	32	-	-	0.3	0.896
CNN	300	32	128	3	0.3	0.918
CNN-LSTM	300	64	256	5	0.5	0.915
CNN-BiLSTM	300	32	256	5	0.3	0.900

The best architecture in our LSTM model builds with 100 units, 32 batch sizes, and 0.5 dropouts. Our LSTM model is then used to evaluate data. The prediction result of our model compares with the actual label of the dataset. The result of aspect classification using evaluation data shows in Table 6.

Table 6. Label Prediction Result on Aspect Classification

Original Text	Preprocessing	Actual Aspect	Predicted Label
<i>Acnya ga dingin trs airnya juga netes trus ke lantai</i> (The AC isn't cool n some water keep continue fall to the floor)	<i>ac tidak_dingin air menetes lantai</i> (ac not_cool waterfall down floor)	room	room
<i>Resepsionisnya baik banget, helpful, proses cekin cekout lumayan cepett</i> (The receptionist was very nice, helpful, and the check-in check-out process is quite fastt)	<i>resepsionis baik helpful proses checkin checkout cepat</i> (the receptionist nice helpful checkin checkout process fast)	service	service
<i>Lokasi dekat banget sama malioboro tinggal jalan kaki</i> (The location is near to Malioboro so we just walked there)	<i>lokasi dekat malioboro hanya jalan kaki</i> (location near Malioboro just walked there)	location	location
<i>Air mineralnya ga gratis</i> (The mineral water isn't free)	<i>air mineral tidak_gratis</i> (mineral water not_free)	price	room

Table 6 shows that the “*air mineral tidak gratis* (mineral water not_free)” sentence gets the wrong prediction. The misclassification in the “mineral water not_free” happens because the document consists of the “*air mineral* (mineral water)” sentence that appears 23 times in the room aspect and seven times in the price aspect for the “tidak gratis (not_free)” sentence. The evaluation matrix uses to see the accuracy value and total of misclassification in every aspect. The evaluation matrix of aspect classification that shows in Figure 4.

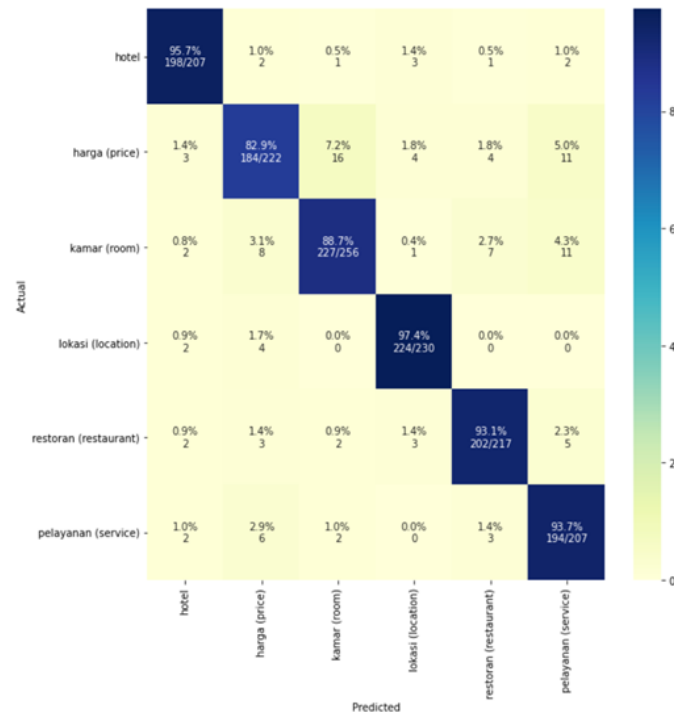


Figure 4. Evaluation Matrix of Aspect Classification

From these results, we can see that the location aspect has the highest accuracy score with a value of 0.974. The lowest accuracy score achieves by the price aspect with a value of 0.829 with a total misclassification of 38 data. The most misclassification aspect belongs to the price aspect that predicts as room aspect. That misclassification occurs because many of the words in the room aspect appear together with the price aspect. For example, in the sentence “*Kita harus bayar lagi buat perlengkapan mandi* (We have to pay more for toiletries)”, “*Harga kamar mahal* (Expensive rooms)”, “*Air mineral ga gratis* (The mineral water isn’t free)”, and others.

3.4 Sentiment Classification

As in aspect classification, sentiment classification also uses some experiments on the architecture and hyperparameters. In this study, we use two scenarios to classify sentiment to get the best result. The first scenario is to classify sentiment in all aspects, and the second one is to classify sentiment in every aspect.

In the first scenario, we train some models with a combination of deep learning algorithms and hyperparameters. From the result, we get the CNN-LSTM model as the highest score with an accuracy value of 0.881543 that shows in Table 7. The architecture of CNN-LSTM as the best model in this study consists of 100 units, 64 batch size, 256 filter size, 3 kernel size, and 0,1 for dropout. This model trains using early stopping for epochs to specify the number of training when the model has stopped improving.

Table 7. Result of Sentiment Classification in All Aspects

Model	Number of Units	Batch Size	Number of Filters	Kernel Size	Dropout	Accuracy
RNN	100	128	-	-	0.5	0.875
LSTM	100	64	-	-	0.1	0.878
GRU	100	32	-	-	0.1	0.880
BiLSTM	100	64	-	-	0.5	0.875
Attention BiLSTM	100	32	-	-	0.3	0.871
CNN	100	64	512	5	0.3	0.872
CNN-LSTM	100	64	256	3	0.1	0.881
CNN-BiLSTM	100	32	16	3	0.5	0.879

The result of the second scenario shows in Table 8. In this scenario, we get the average accuracy score in each model are not too different except in the RNN model. The RNN model has the lowest average accuracy score with an accuracy score of 0.867. Moreover, the RNN model also has the lowest accuracy score in every aspect. Then, the highest score in the scenario classifies sentiment in every aspect gets on the CNN model with a value of 0.904. By comparing the result of the first and second scenarios, the CNN model in the second scenario performs better than the best result in the first scenario. The first scenario gets more misclassifications because some of the words in our dataset labeling with different sentiments in the same words. For example, when the sentences have the “*hanya* (only)” word in price aspect like “*harga hanya seratus ribu* (the price is only one hundred thousand)” get a positive sentiment, and the “only” word in the sentence “*hanya ada satu lift* (there is only one elevator)” with hotel aspect get the negative sentiment.

Table 8. Result of Sentiment Classification in Every Aspect

Model	Price	Hotel	Room	Location	Service	Restaurant	Average
RNN	0.846	0.875	0.875	0.897	0.855	0.855	0.867
LSTM	0.890	0.902	0.907	0.943	0.867	0.869	0.896
GRU	0.887	0.905	0.901	0.953	0.867	0.883	0.899
BiLSTM	0.836	0.888	0.882	0.937	0.893	0.894	0.888
Attention BiLSTM	0.833	0.879	0.879	0.927	0.883	0.883	0.880
CNN	0.880	0.908	0.904	0.950	0.890	0.893	0.904
CNN-LSTM	0.887	0.898	0.901	0.950	0.886	0.900	0.903
CNN-BiLSTM	0.868	0.898	0.895	0.940	0.894	0.876	0.895
Average	0.866	0.894	0.893	0.937	0.879	0.882	

From the result of sentiment classification in every aspect, we also get location and price as the highest and lowest average scores, respectively. In the location aspect, we have an average accuracy score of 0.937 and the price aspect with an average score of 0.866. The location aspect gets a high accuracy value because all sentences have “*dekat* (near)” or “*jauh* (far)” in the test data. Meanwhile, in the price aspect, some misclassifications in this study occurred like “*harga tidak wajar* (unreasonable prices)” and “*harga bersahabat* (friendly prices)”. The misclassification occurs because the words “*wajar* (reasonable)” and “*bersahabat* (friendly)” never appear on training data for the price category. The model with the best results on sentiment classification in every aspect shows in Table 9.

Table 9. Model with The Best Results on Sentiment Classification in Every Aspect

Aspect	Best Model	Accuracy	Number of Units	Batch Size	Number of Filters	Kernel Size	Dropout
Price	LSTM	0.890	100	32	-	-	0.3
Hotel	CNN	0.908	100	32	32	3	0.5
Room	LSTM	0.907	100	128	-	-	0.1
Location	GRU	0.953	300	64	-	-	0.3
Service	BiLSTM	0.894	100	64	-	-	0.5
Restaurant	CNN-LSTM	0.900	100	128	128	3	0.3

The model with the best results on sentiment classification using the second scenario shows in Table 9. The result shows that location and price aspects get the highest and lowest accuracy values, respectively. The location aspect gets the best accuracy using the GRU model with a value of 0.953, while the price aspect gets the best accuracy with a value of 0.890 using the LSTM model. Other than that, the hotel, room, service, and restaurant aspects get the best result by using CNN, LSTM, and BiLSTM models, respectively. The label prediction results on sentiment classification in every aspect that shows in Table 10.

Table 10. Label Prediction Result on Sentiment Classification

Original Text	Preprocessing	Aspect	Actual Label	Predicted Label
<i>Acnya ga dingin trs airnya juga netes trus ke lantai</i> (The AC isn't cool n some water keep continue fall to the floor)	<i>ac tidak_dingin air menetes lantai</i> (ac not_cool waterfall down floor)	room	negative	negative
<i>Resepsionisnya baik banget, helpful, proses cekin cekout lumayan cepett</i> (The receptionist was very nice, helpful, and the check-in check-out process is quite fast)	<i>resepsionis baik helpful proses checkin checkout cepat</i> (the receptionist nice helpful checkin checkout process fast)	service	positive	positive

<i>Lokasi dekat banget sama malioboro tinggal jalan kaki</i> (The location is near to Malioboro so we just walked there)	<i>lokasi dekat malioboro hanya jalan kaki</i> (location near Malioboro just walked there)	location	positive	positive
<i>Air mineralnya ga gratis</i> (The mineral water isn't free)	<i>air mineral tidak gratis</i> (mineral water not_free)	price	negative	negative

4. Conclusion

In this study, we propose a comparison of the deep learning model for ABSA. This study performed good results in aspect classification by using 5,387 with price, hotel, room, location, service, and restaurant aspects. From the comparison of deep learning methods, this works found that the accuracy values in each model are not too much different. The Attention BiLSTM gets the lowest score with an accuracy value of 0.896, and the LSTM model has the best result with an accuracy score of 0.926. Our LSTM model consists of 100 units, 32 batch sizes, and 0.5 dropouts. In addition, the evaluation result using the confusion matrix from this model shows that the price and room aspect has the most misclassified aspect. Those aspects have the highest value of misclassified because items in both classes appear frequently.

In sentiment classification, we use two different approaches. The first one is to classify sentiment in all aspects using a comparison of deep learning methods, and the second one is to classify sentiment in every aspect using the average result of accuracy value of deep learning methods. The result of the experiments shows that classify sentiments in every aspect gets a better result than in all aspects. But in the second scenario, the RNN model performs not pretty well. Besides get the lowest accuracy score, the RNN model also has the lowest average accuracy score in every aspect. The best result in sentiment classification in every aspect shows that the CNN model with an accuracy value of 0.904. Our CNN model consists of 100 units, 64 batch sizes, 256 filters, 3 kernel sizes, and 0.1 dropouts. The number of misclassifications that appear more in the classifies sentiments in all aspects is because some documents have several words with a different label.

References

- [1] T. Tran, H. Ba, and V. N. Huynh, *Measuring hotel review sentiment: An aspect-based sentiment analysis approach*, vol. 11471 LNAI. Springer International Publishing, 2019. https://doi.org/10.1007/978-3-030-14815-7_33
- [2] M. Ady and D. Quadri-Felitti, "Consumer research identifies how to present travel review content for more bookings," *TrustYou*, 2015.
- [3] P. Phillips, S. Barnes, K. Zigan, and R. Schegg, "Understanding the Impact of Online Reviews on Hotel Performance: An Empirical Analysis," *J. Travel Res.*, vol. 56, no. 2, pp. 235–249, 2017. <https://doi.org/10.1177%2F0047287516636481>
- [4] 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*, vol. 11341 LNCS. Springer International Publishing, 2018. https://doi.org/10.1007/978-3-030-04284-4_13
- [5] B. Jang, M. Kim, G. Harerimana, S. U. Kang, and J. W. Kim, "Bi-LSTM model to increase accuracy in text classification: Combining word2vec CNN and attention mechanism," *Appl. Sci.*, vol. 10, no. 17, 2020. <https://doi.org/10.3390/app10175841>
- [6] M. Afzaal, M. Usman, and A. Fong, "Predictive aspect-based sentiment classification of online tourist reviews," *J. Inf. Sci.*, vol. 45, no. 3, pp. 341–363, 2019. <https://doi.org/10.1177%2F0165551518789872>
- [7] H. Peng, Y. Ma, Y. Li, and E. Cambria, "Learning multi-grained aspect target sequence for Chinese sentiment analysis," *Knowledge-Based Syst.*, vol. 148, pp. 167–176, 2018. <https://doi.org/10.1016/j.knosys.2018.02.034>
- [8] D. Ekawati and M. L. Khodra, "Aspect-based sentiment analysis for Indonesian restaurant reviews," *Proc. - 2017 Int. Conf. Adv. Informatics Concepts, Theory Appl. ICAICTA 2017*, 2017. <https://doi.org/10.1109/ICAICTA.2017.8090963>
- [9] S. Wu, Y. Xu, F. Wu, Z. Yuan, Y. Huang, and X. Li, "Aspect-based sentiment analysis via fusing multiple sources of textual knowledge," *Knowledge-Based Syst.*, vol. 183, p. 104868, 2019. <https://doi.org/10.1016/j.knosys.2019.104868>
- [10] S. Gu, L. Zhang, Y. Hou, and Y. Song, "A position-aware bidirectional attention network for aspect-level sentiment analysis," *Proc. 27th Int. Conf. Comput. Linguist.*, pp. 774–784, 2018.
- [11] 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, 2019. <https://doi.org/10.1007/s13042-018-0799-4>
- [12] Y. Luo and X. Xu, "Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: A case study of yelp," *Sustain.*, vol. 11, no. 19, 2019. <https://doi.org/10.3390/su11195254>
- [13] Y. Setiowati, "Service Extraction and Sentiment Analysis to Indicate Hotel Service Quality in Yogyakarta based on User Opinion," *2018 Int. Semin. Res. Inf. Technol. Intell. Syst.*, pp. 427–432, 2016. <https://doi.org/10.1109/ISRITI.2018.8864269>
- [14] L. Vinet and A. Zhedanov, "A 'missing' family of classical orthogonal polynomials," *J. Phys. A Math. Theor.*, vol. 44, no. 8, pp. 329–334, 2011. <https://doi.org/10.1088/1751-8113/44/8/085201>
- [15] J. Thanaki, *Python Natural Language Processing*. 2017.
- [16] P. Prameswari, I. Surjandari, and E. Laoh, "Opinion mining from online reviews in Bali tourist area," *Proceeding - 2017 3rd Int. Conf. Sci. Inf. Technol. Theory Appl. IT Educ. Ind. Soc. Big Data Era, ICSITech 2017*, vol. 2018-Janua, pp. 226–230, 2017. <https://doi.org/10.1109/ICSITech.2017.8257115>
- [17] S. Dolnicar and T. Otter, "Which Hotel Attributes Matter? A review of previous and a framework for future research," *Preterm Birth Prev. Manag.*, pp. 270–273, 2010.
- [18] A. F. Hidayatullah, S. Cahyaningtyas, and R. D. Pamungkas, "Attention-based CNN-BiLSTM for Dialect Identification on Javanese Text," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 4, pp. 317–324, 2020. <https://doi.org/10.22219/kinetik.v5i4.1121>
- [19] SimilarWeb, "Top Apps Ranking."
- [20] H. Jangid, S. Singhal, R. R. Shah, and R. Zimmermann, "Aspect-Based Financial Sentiment Analysis using Deep Learning," pp. 1961–1966, 2018. <https://doi.org/10.1145/3184558.3191827>

- [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. <https://doi.org/10.1007/s10462-019-09794-5>
- [22] A. Mittal, "Understanding RNN and LSTM," towards data science, 2019.
- [23] S. Kostadinov, "Understanding GRU Networks," towards data science, 2017.
- [24] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019. <https://doi.org/10.1016/j.neucom.2019.01.078>
- [25] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," *EMNLP 2017 - Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 452–461, 2017. <http://dx.doi.org/10.18653/v1/D17-1047>
- [26] M. T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *Conf. Proc. - EMNLP 2015 Conf. Empir. Methods Nat. Lang. Process.*, pp. 1412–1421, 2015. <http://dx.doi.org/10.18653/v1/D15-1166>
- [27] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
- [28] J. Xie, B. Chen, X. Gu, F. Liang, and X. Xu, "Self-Attention-Based BiLSTM Model for Short Text Fine-Grained Sentiment Classification," *IEEE Access*, vol. 7, pp. 180558–180570, 2019. <https://doi.org/10.1109/ACCESS.2019.2957510>
- [29] P. Zhou *et al.*, "Attention-based bidirectional long short-term memory networks for relation classification," *54th Annu. Meet. Assoc. Comput. Linguist. ACL 2016 - Short Pap.*, pp. 207–212, 2016. <http://dx.doi.org/10.18653/v1/P16-2034>
- [30] Y. Kim, "Convolutional neural networks for sentence classification," *EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, pp. 1746–1751, 2014. <http://dx.doi.org/10.3115/v1/D14-1181>
- [31] C. L. S.-T. M. Networks, "CNN Long Short-Term Memory Networks," *Machine Learning Mastery*, 2019.
- [32] M. Abdullah, M. Hadzikadicy, and S. Shaikhz, "SEDAT: Sentiment and Emotion Detection in Arabic Text Using CNN-LSTM Deep Learning," *Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018*, pp. 835–840, 2019. <https://doi.org/10.1109/ICMLA.2018.00134>
- [33] Z. Rajabi, O. Uzuner, and A. Shehu, "A Multi-channel BiLSTM-CNN model for multilabel emotion classification of informal text," *Proc. - 14th IEEE Int. Conf. Semant. Comput. ICSC 2020*, pp. 303–306, 2020. <https://doi.org/10.1109/ICSC.2020.00060>
- [34] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, "A CNN-BiLSTM Model for Document-Level Sentiment Analysis," *Mach. Learn. Knowl. Extr.*, vol. 1, no. 3, pp. 832–847, 2019. <https://doi.org/10.3390/make1030048>