Received: April 26, 2022. Revised: August 1, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Aspect-Based Sentiment Analysis for Hotel Review Using LDA, Semantic Similarity, and BERT

Pulung Nurtantio Andono[1] Sunardi[2]* Raden Arief Nugroho[2] Budi Harjo[1]

[1] *Department of Informatics, Universitas Dian Nuswantoro, Semarang, Indonesia*
[2] *Department of English, Universitas Dian Nuswantoro, Semarang, Indonesia*
* Corresponding author's Email: sunardi@dsn.dinus.ac.id

Abstract: Hotel review is frequently used as a main input in sentiment analysis process. It aims at helping travellers easily find more accurate information about hotel aspects in selecting hotels for their journeys. Based on the review datasets, hotel service organizers may evaluate guests' responses towards the services provided by the hotels. Hotel organizers, in turn, may also know the hotel aspects which need improvement for the next experiences. The common problems are because the processed data do not focus on small scale so that wrong selection of terms from review document frequently appears. The problem that often arises is that the amount of data that is processed is not limited to a small scale, so there are often errors in taking terms from a review document. Meanwhile, these terms are the main input source used for the assessment of aspect categorization and aspect-based sentiment analysis. So, we need aspect categorization and aspect-based sentiment analysis methods that can work automatically on a large scale with good accuracy results. In this study, first, the results of the pre-processing were processed using TF-ICF to obtain terms from reviews based on aspect keyword variables in each hotel aspect category. Next, LDA was used to get the hidden topic of each term. The aim was to obtain better terms accuracy results. Then, the aspect categorization process was carried out using BERT embedding and semantic similarity with the aim of obtaining more significant differences in similarity results in each aspect category so that the determination of aspect categories from a review could be more accurate. The results of the aspect extraction obtained an evaluation of the aspect categorization for each precision 0.86, recall 0.92, and f1-measure 0.89. Furthermore, BERT sentiment analysis method is used in the aspect-based sentiment analysis process. Finally, the evaluation result of aspect-based sentiment analysis obtained for each precision, recall, and f-1 measure are 0.96, 0.98, and 0.97.

Keywords: Aspect categorization, LDA, TF-ICF, BERT embedding, Semantic similarity, Aspect-based sentiment analysis, BERT sentiment analysis.

1. Introduction

The development of nowadays digital era enables the booking of traveling needs such as train ticket, flight ticket, tour ticket, hotel reservation, restaurant reservation, and similar facilities via online travel agent website [1]. Travel website also provides features to give review on the services in order that users have best advice in choosing the needed facilities based on the review results from previous users. The results of the facilities review can also be used as supporting data for the services providers to evaluate the experience responses and feedbacks on the given services from the previous customers. In information technology field such an evaluation is called sentiment analysis. Nowadays hotel review is frequently used as a main input in sentiment analysis process. It aims at helping travellers easily find more accurate information about hotel aspects in selecting hotels for their journeys.

Sentiment analysis is a computational technique for automatic extraction of customer opinions regarding a product [2]. The development of sentiment analysis itself has reached the stage of aspect-based sentiment analysis. Aspect-based sentiment analysis (ABSA) is a type of sentiment analysis that can produce sentiment assessments on predetermined aspects [3]. ABSA can infer data about the polarity of a sentiment based on the aspect category or target entity in the text [4]. One of the widely used ABSA implementations is for hotel review ratings [5, 6, 7].

In ABSA, in order to make the assessment of hotel reviews, the hotel review dataset is used to find out what aspects need to be improved [8]. From the dataset from the review, hotel service managers can assess consumer reactions to the services provided by the hotel so that they can evaluate these services. Furthermore, hotel managers can also understand what needs to be improved for future experiences. Therefore, accurate determination of the aspect and opinion of word is necessary, so that there is no mistake in determining the aspect category and sentiment polarity of a review.

In previous studies, various ABSA methods and techniques for hotel reviews have been carried out to work on cases of term extraction, aspect categorization, and sentiment analysis contained in it. Research related to ABSA for hotel reviews [6] has used probabilistic latent semantic analysis [9] to take the aspect of term in hotel review. The research still works on small-scale data and the type of sentence that is involved in this work is also a simple sentence. An example of the data includes "*a clean swimming pool at Manhattan*". The sentence datum that is done is only a simple sentence which contains one aspect term and one opinion term. This study has not done the extraction of hotel reviews on large-scale data and has not done a hotel review consisting of several sentences. In line with the purpose of using the method in this study, latent dirichlet allocation (LDA) [10] is one of the most widely used unsupervised learning methods in topic modelling to help determine the topic of a text through hidden topics from a document. [11, 12]. In previous hotel review research [7], LDA is used to extract aspect and opinion terms from hotel reviews for the ABSA proses process. Similar to the previous research [6], research [7] also uses LDA. It works on simple sentence cases, for example "*I can hear bass sound*" and does not work on extracting hotel reviews on large-scale data and review cases containing more than one sentence. The use of LDA itself depends on the number of hidden topics used. If in one review there are several long sentences, this can certainly affect the accuracy of determining LDA hidden topics, so that aspect acquisition and opinion terms become inaccurate. These inaccurate outcomes can also result in incorrect measurements in the word similarity process.

Word similarity process is a word extraction process to get the word similarity of aspect term and category. The most frequently used word similarity is semantic similarity [13, 14, 15]. Semantic similarity is token-dependent, which is generated to be used as input, namely terms. Moreover, if the data are large and have multiple sentences, they will greatly affect the selection of accurate terms as needed. Therefore, it is necessary to develop word similarity procedures that can increase the accuracy of this process. Other studies [16] have worked on novel token-based text extraction techniques, one of them is BERT. BERT [17, 18] BERT, one of the methods produced, can work automatically on large-scale data. In the word similarity process, BERT has provided a very significant development. However, it needs to be developed if the generated tokens have close word similarity values among each category class; By doing so, it can lead to zero errors in the categorization process.

The output of word similarity, in turn, is used as the input of ABSA process. ABSA process is used to evaluate whether a review is positive or negative based on the word opinion in a sentence. There are many methods to classify positive and negative reviews in ABSA process. Support vector machine (SVM) is one of the mostly used methods to classify aspect-based sentiments [19]. However, the use of SVM needs some improvements because of its process on the small-scale data.

This study focuses on ABSA for hotel review to improve the novelty of previous studies. This study works on a case review that is not only in the form of small-scale data but also on large-scale data. The proposed method works automatically on the aspect categorization and sentiment analysis process for reviews containing simple sentences and more complex sentences, for example, "*wonderful location lots of extra value, great location lots extras, water bottles room complimentary …*". Firstly, review pre-processing needs to be carried out in order that the taken raw data are ready to be processed in the next process, that is, terms extraction using TF-ICF. Then, topic modelling process using LDA is used to obtain hidden topics based on the variables of the keywords of each hotel aspect category in the review. Subsequently, the terms generated from the LDA are expanded to improve the accuracy of terms retrieval using BERT embedding. In aspect categorization process, semantic similarity is used to measure the similarity between aspect terms and categories, while BERT sentiment analysis is used in sentiment analysis. At last, evaluation of aspect categorization process and aspect-based sentiment analysis is carried out to find the score of best performance based on precision, recall, f1-measure, and accuracy.

Received: April 26, 2022. Revised: August 1, 2022.

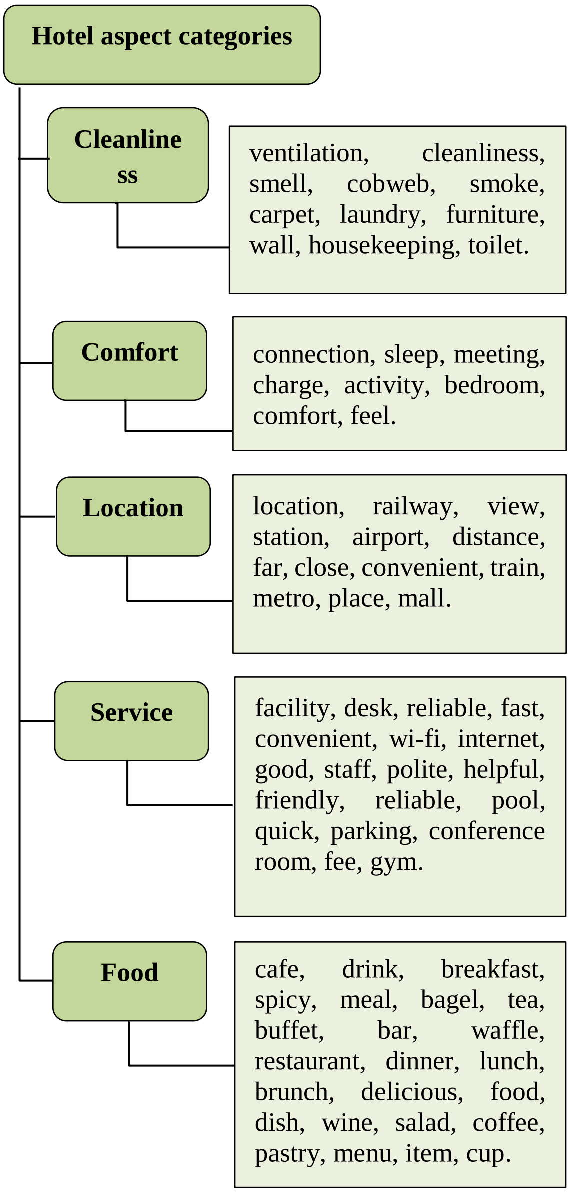

Figure. 1 Keyword terms for hotel aspect categories

2. Related theory

Several theories related to this study are explained in this section.

2.1 Data pre-processing

The pre-processing stage for text extraction processing is carried out to process raw data so that it is ready to be processed based on data requirements that will be used as input for further analysis processes. Generally, the pre-processing stages [20] are 1) case folding, 2) filtering, 3) normalization, 4) stop word removal, 5) stemming, and 6) tokenizing.

2.2 Keyword for hotel aspect

Five aspects have been found out from the previous studies [7 6], namely location, food, cleanliness, comfort, and service. Fig. 1 shows keyword terms for the five aspects.

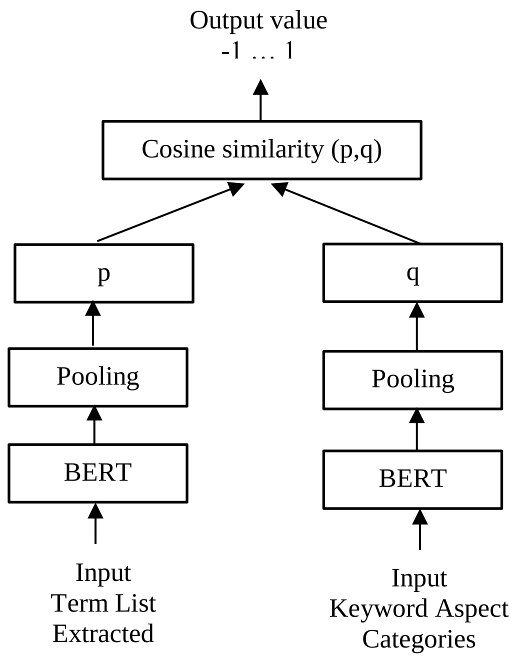

Figure. 2 BERT architecture to compute word similarity

2.3 TF - ICF

Term frequency - inverse cluster frequency (TF - ICF) [21, 22] is a term weighting scheme that used as a method for clustering process by calculating the term weight of a word against document of each cluster. The ICF equation is shown in Eq. (1) and TF-ICF equation is shown in Eq. (2).

$$ICF_x = 1 + \log \frac{c}{cf_x} \tag{1}$$

$$TF\text{-}ICF_x = TF_{yx} \times ICF_x \tag{2}$$

Where ICF_x is the Inverse Cluster Frequency value of word x, CF_x is cluster numbers containing word x, and c is the exist cluster numbers, TF_{yx} is word x frequency in cluster y, $TF - ICF_x$ is weighting value of word i from product result of TF_{yx} and ICF_x.

2.4 LDA

Latent dirichlet allocation (LDA) method has been widely used in topic modelling [23]. It is also better than latent semantic analysis (LSA) and probabilistic latent semantic analysis (PLSA) [24]. The equation of LDA method can be defined as [7]:

$$p(w, z/\alpha, \beta) = p(w/\alpha, \beta)\, p(z/\alpha) \tag{3}$$

Received: April 26, 2022. Revised: August 1, 2022.

		Actual Values	
		Positive	**Negative**
Predicted Values	**Positive**	*True Positive (TP)*	*False Negative (FN)*
	Negative	*False Positive (FP)*	*True Negative (TN)*

Figure. 3 Confusion matrix

where α is first model parameter, β is second model parameter, w is word target in document, z is topic in document, and $p(z/\alpha)$ is topic z probability.

2.5 Semantic similarity

Semantic similarity is a measure for comparing the semantic similarity of words and phrases that establish values based on semantic relationships. Semantic similarity metrics are used to compare words and terms in natural language texts and are calculated to compute the similarity between concepts [25]. The semantic relation with each term in the LDA extracted term and the keyword phrase, as well as the expansion results. Applying cosine similarity to estimate the score of similarity between words (w_i, wj) is based on the calculation of the number of similar words, where the Cosine similarity is calculated using the word vector [26]. The equation of semantic similarity formula [6] is described as follows:

$$Similarity = \frac{\sum_{m-1}^{k} w_i{}^m wj^m}{\sqrt{\sum_{m-1}^{k}(w_i{}^m)^2}\sqrt{\sum_{m-1}^{k} wj^m)^2}} \quad (4)$$

Cosine similarity [26] determines the degree of similarity between sentences 1 (S1) and 2 (S2) by counting the number of similar terms in both. Word vectors are used to measure the cosine similarity with the formula equation as follows:

$$Cosine(S1, S2) = \frac{\sum_{i=1}^{k} S1i\, S2i}{\sqrt{\sum_{i=1}^{k} S1i^2}\sqrt{\sum_{i=1}^{k} S2i^2}} \quad (5)$$

2.6 BERT

Bidirectional encoder representations from transformers (BERT) [17] is used for language modelling based on bidirectional transformer training (attention model). BERT uses attention mechanism which gets the real context of text and has two important parts: encoder (to get text input) and decoder (to produce output or task prediction).

2.6.1. Word embedding using BERT

BERT architecture to compute word similarity score [18] can be illustrated in Fig 2.

2.6.2. Sentiment analysis using BERT

BERT method is used to sentiment analysis process. The following are sentiment analysis steps using BERT method [17]:

1. Data pre-processing
2. Splitting data train and test
3. Tokenization, including token, segment, and position embedding
4. Encoding
5. Set up model
6. Set up BERT pre-trained Model
7. Create data loaders
8. Set up optimizer and scheduler
9. Define performance matrix
10. Evaluation

2.7 Evaluation

In this study, the evaluation process is used to measure accuracy score of aspect categorization and sentiment analysis. The evaluation of work performance uses confusion matrix approach [27] to calculate precision, recall, and F1-measure. The confusion matrix can be seen in Fig. 3.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (9)$$

3. Research method

This study started with pre-processing the dataset as the raw input data, to prepare relevant data information. The relevant data were subsequently processed in the aspect categorization step using LDA method to determine the hidden topic and TF-ICF to get the determined aspect term numbers. The next semantic similarity was done using BERT embedding and cosine similarity. The last process was aspect-based sentiment analysis using BERT

 DOI: 10.22266/ijies2022.1031.21

Received: April 26, 2022. Revised: August 1, 2022.

Table 1. Example of data collection

No.	Review
1.	favorite hotel work responsibilities taken seattle 4-6 times year years stayed hotel andra dozen times 2-5 nights stay …
2.	wonderful location lots extra value, great location lots extras, water bottles room complimentary …
3.	nice hotel expensive parking got good deal stay hotel anniversary, arrived late evening took advice previous reviews did valet parking …
4.	ok nothing special charge diamond member hilton decided chain shot 20th anniversary seattle …
5.	nice rooms not 4* experience hotel monaco seattle good hotel n't 4* level. positive large bathroom mediterranean suite …

Table 2. Result of pre-processing sentence[1]

ID	Document Review	Pre-processing Result
1	wonderful location lots extra value, great location lots extras, water bottles room complimentary, complimentary wine reception 5 6 pm night..great way meet guest parts world, like europeans staying complete continental breakfast helpful staff restaurants transportation, told staff worked 20 years, quite testament hotel/owners, small room hotel totally re-furbished 04 great bed linens small clean really need, 1/2 block union square street car line not ask 129.00 night, awesome,	[['wonderful', 'location', 'lot', 'extra', 'value', 'great', 'location', 'lot', 'extra', 'water', 'bottle', 'room', 'complimentary', 'complimentary', 'wine', 'reception', '5', '6', 'pm', 'nightgreat', 'way', 'meet', 'guest', 'part', 'world', 'like', 'european', 'staying', 'complete', 'continental', 'breakfast', 'helpful', 'staff', 'restaurant', 'transportation', 'told', 'staff', 'worked', '20', 'year', 'quite', 'testament', 'hotelowners', 'small', 'room', 'hotel', 'totally', 'refurbished', '04', 'great', 'bed', 'linen', 'small', 'clean', 'really', 'need', '12', 'block', 'union', 'square', 'street', 'car', 'line', 'ask', '12900', 'night', 'awesome']]

method to get positive and negative sentiment values from the review sentences used as the dataset.

3.1 Dataset

The dataset used in this study describes the source of the data obtained in the form of reviews from various hotel customers with datasets + crawling data review using WebHarvy on tripadvisor.com website. The dataset contains the review sentences given by the customers as exemplified in Table 1.

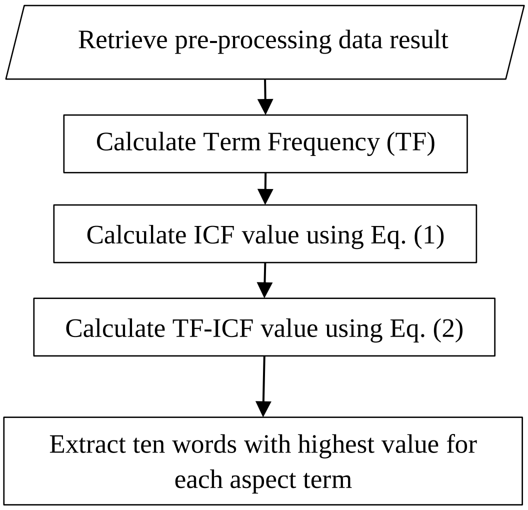

Figure. 4 TF-ICF process

3.2 Pre-processing

In this study, the pre-processing stage was done using the following steps: 1) text to lower case conversion, 2) number removal, 3) Greek character removal, 4) stop words removal, 5) Lemmatization, and 6) Stemming. Performing the steps in doing text pre-processing [28] can be shown in Table 2 as follows:

3.3 Aspect term extraction

The aspect term extraction was done using TF-ICF. TF-ICF was also used to determine aspect term list and to take ten term lists with highest values based on the calculation using TF-ICF formula in Eq. (1) and (2). The flow chart of TF-ICF process can be shown in Fig. 4.

3.4 Topic modelling using latent dirichlet allocation (LDA)

Results obtained from the LDA process are hidden topics and the frequency value of the possibility of words appearing from the document. Bag-of-word LDA corpus is used to calculate the occurrence of each word in the document using. Then build a training LDA model that calculates the possible proximity of related topics. Then, LDA produces hidden topics, which are key words or phrases that appear frequently in documents. The hidden topic extraction results from each term list document using this approach are as follows:

The results of the LDA are hidden topics and terms on documents that can be seen in Table 3. Each aspect as a whole is calculated using TF-ICF and Semantic Similarity. The results of semantic similarity to categorize each review on the five specified hotel aspects are carried out

Table 3. Hidden topic results of LDA

Term Topic	Hidden Topic
Term List Topic 1	0.037*"hotel" + 0.030*"time" + 0.022*"restaurant" + 0.019*"excellent" + 0.019*"location"
Term List Topic 2	0.030*"location" + 0.029*"room" + 0.026*"staff" + 0.025*"great" + 0.025*"extra"
Term List Topic 3	0.050*"nice" + 0.027*"room" + 0.027*"parking" + 0.022*"stay" + 0.019*"night"
Term List Topic 4	0.022*"room" + 0.022*"hotel" + 0.015*"suite" + 0.015*"website" + 0.013*"desk"
Term List Topic 5	0.030*"room" + 0.018*"night" + 0.017*"stay" + 0.016*"wakeup" + 0.014*"great"

Input: Aspect term list, Aspect keyword
Output: Score similarity
Start
1. Tokenize aspect term list;
2. Determine where tokenizer has word_ids for aspect word;
3. Generate hidden state of aspect word;
4. Extract the token after corresponding to aspect word;
5. Average.

End

Figure. 5 BERT embedding algorithm

Input: ABSA pre-processing result
Output: Sentiment polarity
Start
1. Splitting data train and test;
2. Creating input sequence;
3. Configuring model and fine-tuning;
4. Making prediction of sentences.

End

Figure. 6 ABSA using BERT algorithm

Table 4. Result of AC1

ID	Aspect 1	Aspect 2	Aspect 3	Aspect 4	Aspect 5
	Clean.	Comfort	Location	Service	Food
0	0.1179	0.4811	0.2520	0.3195	0.2981
1	0.1179	0.3849	0.2520	0.4108	0.3727
2	0.1179	0.4811	0.3150	0.4564	0.3354
3	0.1179	0.3849	0.3780	0.4564	0.3727
4	0.1179	0.4811	0.2520	0.3195	0.3727

using cosine similarity calculations.

3.5 BERT embedding

Document expansion in aspect terms from LDA is subsequently followed by similarity matching process using BERT embedding to improve the accuracy of term selection which will be processed in the next step. The BERT embedding steps in this study can be shown in Fig. 5.

3.6 Aspect categorization

Aspect categorization process was done using semantic similarity method to categorize aspect terms into five hotel aspects. The main concept of this step was to extract hidden topic data in the document using latent dirichlet allocation (LDA). The results of LDA were hidden topic data which were subsequently calculated to know their word similarities towards five aspect categories using semantic similarity.

In aspect categorization (AC) step, three methods were used to determine the best performance of aspect categorization: aspect categorization 1 (AC1) was done using LDA, TF-ICF 20% and semantic similarity; aspect categorization 2 (AC2) was done using LDA, TF-ICF 100% and semantic similarity; aspect categorization 3 (AC3) was done using LDA, TF-ICF 100%, BERT embedding, and semantic similarity.

3.7 Aspect based sentiment analysis

Pre-processing for ABSA

In this process, a description of the resulting data from the categorization aspect term list for each aspect that has been categorized was collected based on the relevant aspect. Category of each term on each topic was based on the same aspects. After that, the sentiment analysis was determined for all data related to the term list based on five aspects. The labelling process was carried out on each data regarding positive and negative sentiments in all aspects.

ABSA

In this stage, a term list for each aspect that has been categorized was collected based on the relevant aspect. After that, sentiment analysis was determined on all data related to the term list in each of the five aspects. The labelling process was carried out on each data regarding positive and negative sentiments in all aspects. Sentiment analysis was determined based on the relevant aspect, for example, data containing sentiment regarding the cleanliness aspect. Classification was carried out based on the data that has been carried out by training that has been categorized as cleanliness aspect. Then, the sentiments of the data containing the term list were classified based on the aspects. The ABSA steps using BERT method can be shown in Fig. 6.

Table 5. Result of AC2

ID	Aspect 1	Aspect 2	Aspect 3	Aspect 4	Aspect 5
	Clean.	Comfort	Location	Service	Food
0	0.0870	0.4975	0.1861	0.4045	0.3853
1	0.0913	0.3727	0.2440	0.3182	0.3464
2	0.0913	0.4472	0.2928	0.4243	0.3753
3	0.0913	0.4472	0.3904	0.4243	0.3175
4	0.0870	0.3553	0.2791	0.4045	0.3853

Table 6. Comparison of AC performances

Approach	Methods	F-1 Measure
AC1	LDA + TF-ICF 20% + Semantic Similarity	0.75
AC2	LDA + TF-ICF 100% + Semantic Similarity	0.81
AC3	LDA + TF-ICF 100% + BERT + Semantic Similarity	0.89

Table 7. Result of AC3

ID	Aspect 1	Aspect 2	Aspect 3	Aspect 4	Aspect 5
	Clean.	Comfort	Location	Service	Food
0	0.8186	0.8936	0.9061	0.8674	0.8848
1	0.8384	0.8834	0.8823	0.8952	0.8644
2	0.8416	0.9114	0.9011	0.9075	0.8719
3	0.8749	0.9111	0.9050	0.8779	0.8833
4	0.8146	0.9149	0.8667	0.8976	0.8717

3.8 Evaluation

To get the best performance, this study applied the following steps: precision, recall, and f-1 measure.

4. Result and analysis

4.1 Aspect categorization result

Aspect categorization (AC) process in this study produced three results from three testing: AC1, AC2, and AC3.

4.1.1. Aspect categorization 1

Table 4 shows the results of aspect categorization 1 (AC1) process using LDA, TF-ICF 20%, and Semantic Similarity. It shows that AC1 can well extract the aspect terms in five aspect categories. However, it also shows that the assessment carried out is not significant enough to distinguish the acquired aspect of words because the value generated in the review with ID [0] to [4] averages below 50% or 0.5. This probably happens because of the wrong selection of the real aspect terms.

4.1.2. Aspect categorization 2

Table 5 shows the results of aspect categorization 2 (AC2) process using LDA, TF-ICF 100%, and Semantic Similarity. It shows that AC2 is better than AC1. This is identified from the assessment in the Aspect1 column which starts to improve, unlike AC1 which produces an Aspect 1 value which is the same Cleanliness as in 5 reviews. However, the average similarity value resulting from the AC2 process is still the same as AC1 which is below 50% or 0.5. This also allows the occurrences of errors in taking the actual aspect terms.

4.1.3. Aspect categorization 3

Table 6 shows comparison of results from AC1, AC2, and AC3. It also shows that aspect categorization 3 (AC3) has the best performance value with F-1 measure 0.89.

Table 7 shows the results of AC3 using LDA, TF-ICF 100%, BERT, and semantic similarity. It also shows that AC3 testing values are better than AC1, and AC2. Firstly, the results of term selection from the hidden topic made with LDA are done using TF-ICF. In AC3, the results of TF-ICF from AC2 with 20% are expanded to 100%. Word embedding using BERT is subsequently added to AC3 testing to get more accurate results of word similarity based on the aspect terms taken in LDA and TF-ICF.

AC3 results are significantly enough as shown in Table 8. It also shows that the results of AC3 testing in each aspect category column are better than AC1 and AC2 in five review sentences. AC3 values show better results of word similarity towards five aspect categories. This contrasts to the results of AC1 and AC2 which still undergo wrong testing as seen from the same values of testing a category aspect. AC3 can be more accurate and varied in selecting and testing aspect terms in the review based on five aspect categories. Table 8 shows the results of AC3 where term list [0] has highest similarity value on Location, term list [1] on Service, term list [2] on Comfort, term list [3] on Comfort, and term list [4] on Comfort. Total scores of evaluations using AC3 method for precision, recall, and f-1 measure are successively 0.86, 0.92, and 0.89.

Table 8 shows comparison of aspect term result between AC1, AC2, and AC3 that is doing by algorithm and aspect term result that determined by expert. In Table 8, IDReview#0 has been assessed by the expert regarding the aspect categorization, namely LOCATION. In IDReview#0, the extraction of aspect terms AC1 and AC2 results in the categorization of the review into COMFORT aspects.

Received: April 26, 2022. Revised: August 1, 2022.

Table 8. Comparison of AC process results

Review		Aspect term result	Aspect term results		
ID	**Pre-processing result**		**AC1**	**AC2**	**AC3**
		Doing by expert	Doing by algorithm	Doing by algorithm	Doing by algorithm
0	'favorite', 'hotel', 'work', 'responsibility', 'taken', 'seattle', '46', 'time', 'year', 'year', 'stayed', 'hotel', 'andra', 'dozen', 'time', '25', 'night', 'stay', 'time', 'period', 'travel', 'extensively', 'domestically', 'internationally', 'favorite', 'hotel', 'apart', 'superb', 'location', 'seam', 'downtown', 'seattle', 'bell', 'town', 'wonderful', 'restaurant', 'location', 'beat', 'view', 'nonconvention', 'style', 'hotel', 'contemporary', 'furnishing', 'comfortable', 'bed', 'efficient', 'workspace', 'easy', 'internet', 'access', 'excellent', 'restaurant', 'accessible', 'lobby', 'breakfast', 'lola', 'outstanding', 'assaggio', 'terrific', 'italian', 'place', 'dinner', 'street', 'excellent', 'lunch', 'dinner', 'dahlia', 'lounge', 'staff', 'particularly', 'friendly', 'responsive', 'happy', 'accommodate', 'reasonable', 'request', 'favor', 'news', 'junkie', 'really', 'value', 'fact', 'ask', 'presto', 'free', 'copy', 'new', 'york', 'time', 'including', 'sunday', 'door', 'morning'	location, restaurant's location, view, beds, workspaces, internet, restaurants, breakfasts, Italian place, lunch, dinner, staff	time, hotel, dinner, Seattle, year, location, restaurant, contemporary	time, hotel, dinner, Seattle, year, location, restaurant, favorite, contemporary, stay, staff, period, street, night, door, view, pace, place, comfort, friendly	hotel, time, location, restaurant, view, comfort, bed, workspace, internet, breakfast, place, lunch, staff, friendly
	Total of aspect categorization result	LOCATION	COMFORT	COMFORT	LOCATION
1	'wonderful', 'location', 'lot', 'extra', 'value', 'great', 'location', 'lot', 'extra', 'water', 'bottle', 'room', 'complimentary', 'complimentary', 'wine', 'reception', '5', '6', 'pm', 'nightgreat', 'way', 'meet', 'guest', 'part', 'world', 'like', 'european', 'staying', 'complete', 'continental', 'breakfast', 'helpful', 'staff', 'restaurant', 'transportation', 'told', 'staff', 'worked', '20', 'year', 'quite', 'testament', 'hotelowners', 'small', 'room', 'hotel', 'totally', 'refurbished', '04', 'great', 'bed', 'linen', 'small', 'clean', 'really', 'need', '12', 'block', 'union', 'square', 'street', 'car', 'line', 'ask', '12900', 'night', 'awesome'	Location, location, water bottles, wine, breakfast, staff, staff, owner, small room hotel, bed linens	room, staff, location, small, lot, extra, complimentary, great, complete, hotel	room, staff, location, small, lot, extra, complimentary, great, complete, hotel, stay, wonderful, like, street, night, block, small, helpful	Location, room, wine, way, breakfast, staff, owner, room, bed, linen
	Total of aspect categorization result	SERVICE	SERVICE	COMFORT	SERVICE

Meanwhile, AC3 result in the categorization of reviews into the LOCATION aspect. The results of aspect terms acquisition can affect the determination of aspect categories, such as the semantic similarity measurement which has been shown in Tables 4, 5, and 6. So, it shows that AC3 can work well because it is able to produce the correct IDReview#0 aspect category as determined by the expert.

In IDReview#1, the extraction of AC1 aspect terms results in the categorization of the review into the SERVICE aspect. AC2 results in the categorization of reviews into COMFORT aspects. Meanwhile, AC3 results in the categorization of reviews into the SERVICE aspect. Tables 4, 5, and 6 have also shown the calculation of semantic similarity of the three AC processes for IDReview#1. Here, AC1 can assess the review because it is in accordance with what has been determined by the expert. However, the average similarity value is still below 50%. Meanwhile, the similarity value generated by AC3 can be above 80% with the highest value ranging from 90%. Therefore, it can be identified that AC3 can work well because it is also able to produce the correct IDReview#1 aspect category as determined by the expert.

4.2 ABSA result

Table 9 shows the evaluation results of ABSA. It shows a comparison between ABSA performance in

Table 9. Result of ABSA

ABSA Performance				
ABSA Approach	**Precision**	**Recall**	**F1-Measure**	**Accuracy**
Reza	0.91	0.96	0.93	-
Dewi	0.93	0.96	0.95	-
Proposed method	**0.96**	**0.98**	**0.97**	**0.97**

this study using BERT and the previous studies. It can be inferred that ABSA method in this study is better than that in the previous studies with precision score 0.96, recall 0.98, f-1 measure 0.97, and accuracy 0.97.

The ABSA results using BERT sentiment analysis are presented in Table 10 which shows that the results of the proposed ABSA method can work better on documents containing several sentences. The process of taking opinion terms is carried out automatically and more accurately to get predictions for positive and negative sentiment categories based on the results of AC3. The results of the aspect categories of each review using AC3 are used as the main input for ABSA to obtain opinion terms. The results of extracted opinion terms are the opinion terms that exist in the review that have a relationship with the aspect terms of aspect category results in AC3.

In IDReview, [0] the review generated aspect category **LOCATION** with aspect terms "location, location, view, place". These terms are then processed to get the word opinion ["superb", "wonderful", "not beat", "terrific"] which finally result into pairing terms [Location: "Superb"], [location: "wonderful"], [view: "not beat"], and [place: "terrific"]. Therefore, it creates **Positive** sentiment polarity.

For the next review in the dataset, which has been pre-processed for the ABSA process, the ABSA process is the same as the ABSA explanation in IDReview review [0].

The extraction result of the aspect category of IDReview [1] is **SERVICE**. In IDReview [1], we can obtain pairing terms [room: "complimentary"], [wine: "complimentary"], [staff: "helpful"], [owner: "testament"], [hotel: "refurbished"], [linen: "great"]. Hence, it creates **Positive** sentiment polarity.

The extraction result of the aspect category of IDReview [2] is **COMFORT**. In IDReview [2], we can obtain [deal: "good"], [room: "clean"], [size: "nice"], [bed: "comfortable"], [pillows: "high"], [bath: "nice"] that creates **Positive** sentiment polarity.

Table 10. ABSA result

ID	**Aspect category extracted**	**Pairing terms**	**Result**
0	Location	[Location: "Superb"], [location: "wonderful"], [view: "not beat"], [place: "terrific"]	Positive
1	Service	[room: "complimentary"], [wine: "complimentary"], [staff: "helpful"], [owner: "testament"], [hotel: "refurbished"], [linen: "great"]	Positive
2	Comfort	[deal: "good"], [room: "clean"], [size: "nice"], [bed: "comfortable"], [pillows: "high"], [bath: "nice"]	Positive
3	Comfort	[charge: "nothing special"] [paid: "extra"] [bedroom: "suite"] [room: "standard"] [advertising: "false"] [email: "failure"] [champagne: "no"] [pillow: "no"] [room: "great"] [beds: "comfortable"] [ac: "not good"] [screen: "shine"]	Negative
4	Comfort	[room: "not 4"], [bathroom: "large"] [pillow: "comfortable"] [ac: "malfunctioned"] [desk: "disorganized"] [tv: "hard"] [dock: "non functioning"] [tub: "great"] [shower: "nice"] [accessories: "no"] [tv: "far"] [dock: "broken"] [box: "not work"] [room: surprised] [amenities: "not"]	Negative

The extraction result of the aspect category of IDReview [3] is **COMFORT**. In IDReview [3], we can obtain pairing terms [charge: "nothing special"],

Table 11. ABSA Result for each category

Results of Sentiment Based Each Aspect		
Aspect	**Sentiment**	**Results Evaluation (Percentage)**
Cleanliness	Positive	10.82
	Negative	1.14
Comfort	Positive	24.42
	Negative	2.67
Location	Positive	10.17
	Negative	1.28
Service	Positive	33.34
	Negative	6.20
Food	Positive	9.08
	Negative	0.88
Total		**100.00**

[paid: “extra”], [bedroom: “suite”], [room: “standard”], [advertising: “false”], [email: “failure”], [champagne: “no”], [pillow: “no”], [room: “great”], [beds: “comfortable”], [ac: “not good”] that creates **Negative** sentiment polarity.

The extraction result of the aspect category of IDReview [4] is **COMFORT**. In IDReview [4], we are able to obtain pairing terms [room: “not 4“], [bathroom: “large”], [pillow: “comfortable”], [ac: “malfunctioned”], [desk: “disorganized”], [tv: “hard”], [dock: “non functioning”], [tub: “great”], [shower: “nice”], [accessories: “no”], [tv: “far”], [dock: “broken”], [box: “not work”], [room: surprised], [amenities: “not”] that creates **Negative** sentiment polarity.

Based on the results of the sentiment analysis of the 5 reviews, it can be concluded that the proposed ABSA method can work well based on the aspect category results obtained previously. The proposed ABSA method can take all opinion terms based on the aspect category results.

Table 11 shows the ABSA results for the aspect categories of cleanliness, comfort, location, service, and food. It also shows that the aspect category of Service has highest percentage of positive sentiment with 33.34%. The other positive sentiment percentages are comfort with 24.42%, cleanliness with 10.82%, location with 10.17%, and food with 9.08%. Meanwhile, the highest percentage of negative sentiment is service with 6.20%. The other negative sentiment percentages are comfort with 2.67%, location with 1.28%, cleanliness with 1.14%, and food with 0.88%.

5. Conclusion and future work

This study proposes ABSA for hotel review using LDA, semantic similarity, and BERT.

At the initial stage, the pre-processing results are processed using TF-ICF to obtain the overall terms. Then, the topic modeling process is carried out using LDA to get the aspect terms based on hidden topics related to the 5 categories of hotel aspects that have been determined. Furthermore, the aspect term expansion is carried out using BERT embedding to increase the accuracy of aspect terms acquisition generated from the LDA process. Here, we use BERT embedding to add an expansion of words aspects and opinions contained in the review based on the terms of BERT document. Furthermore, the results of LDA and word expansion from BERT are combined to retrieve word similarity in 5 aspect categories using semantic similarity. The result of the best aspect categorization performance is AC3 approach with F1-measure of 0.89. Meanwhile, in the aspect-based sentiment analysis process, BERT method is utilized. The results of the aspect-based sentiment analysis performance obtain the F1-measure value of 0.97. These results are better than the 2 previous studies that have been conducted on hotel review ratings.

In this study, we are able to carry out large-scale data extraction containing reviews that not only consist of simple sentences but also more than that. We are able to do aspect categorization and sentiment analysis automatically using the method we have proposed in this research. However, we have not been able to discuss the case of sentence extraction for words with implicit aspect and opinion. Hence, we hope that further research can work on this issue. Other aspect categorization methods can be developed to improve the accuracy of word extraction aspects and opinions, especially: 1) in the case of implicit aspect, 2) in the case of implicit opinion, and 3) in the case of implicit aspect and opinion. In addition, it is necessary to develop a method to obtain aspect terms in the form of word phrases which can extract multi-categories aspects contained in a review. The purpose of this development is to not only produce a total assessment of the aspect categorization process, but also to produce more specific data related to the assessment of aspect-based sentiment analysis from each review.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

This research can work well and successfully because of the following research contributions: Conceptualization by Pulung Nurtantio Andono, Sunardi, Raden Arief Nugroho, and Budi Harjo; methodology by Sunardi, Raden Arief Nugroho;

software usage by Pulung Nurtantio Andono and Budi Harjo; validation, formal analysis, investigation, resources by Pulung Nurtantio Andono, Sunardi, Raden Arief Nugroho, and Budi Harjo; data curation by Budi Harjo; writing-original draft preparation and writing-review and editing by Sunardi and Raden Arief Nugroho; visualization by Budi Harjo; supervision by Pulung Nurtantio Andono; and project administration and funding acquisition by Sunardi.

Acknowledgments

This research is supported by Universitas Dian Nuswantoro and Directorate of Research and Community Service, Directorate General of Research and Development Strengthening of Ministry of Research, Technology and Higher Education of the Republic of Indonesia.

References

[1] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, “Sentiment Analysis Using Word2vec and Long Short-Term Memory (LSTM) for Indonesian Hotel Reviews”, *Procedia Comput. Sci.*, Vol. 179, pp. 728–735, 2021, doi: 10.1016/J.PROCS.2021.01.061.

[2] B. S. Rintyarna, R. Sarno, and C. Fatichah, “Semantic features for optimizing supervised approach of sentiment analysis on product reviews”, *Computers*, Vol. 8, No. 55, pp. 1-16, 2019, doi: 10.3390/computers8030055.

[3] S. Cahyaningtyas, D. H. Fudholi, and A. F. Hidayatullah, “Deep Learning for Aspect-Based Sentiment Analysis on Indonesian Hotels Reviews”, *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, Vol. 6, No. 3, pp. 239-248, 2021, doi: 10.22219/kinetik.v6i3.1300.

[4] W. Xue and T. Li, “Aspect based sentiment analysis with gated convolutional networks”, In: *Proc. of ACL 2018 - 56th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. Long Pap.*, Vol. 1, pp. 2514–2523, 2018, doi: 10.18653/v1/p18-1234.

[5] I. Perikos, K. Kovas, F. Grivokostopoulou, and I. Hatzilygeroudis, “A system for aspect-based opinion mining of hotel reviews”, In: *Proc. of 13th Int. Conf. Web Inf. Syst. Technol.*, No. Webist, pp. 388–394, 2017, doi: 10.5220/0006377103880394.

[6] D. Khotimah and R. Sarno, “Sentiment Analysis of Hotel Aspect Using Probabilistic Latent Semantic Analysis, Word Embedding and LSTM”, *Int. J. Intell. Eng. Syst.*, Vol. 12, No. 4, pp. 275–290, 2019, doi: 10.22266/ijies2019.0831.26.

[7] R. Priyantina and R. Sarno, “Sentiment Analysis of Hotel Reviews Using Latent Dirichlet Allocation, Semantic Similarity and LSTM”, *Int. J. Intell. Eng. Syst.*, Vol. 12, No. 4, pp. 142–155, 2019, doi: 10.22266/ijies2019.0831.14.

[8] R. Annisa, I. Surjandari, and Zulkarnain, “Opinion mining on mandalika hotel reviews using latent dirichlet allocation”, *Procedia Comput. Sci.*, Vol. 161, pp. 739–746, 2019, doi: 10.1016/j.procs.2019.11.178.

[9] T. Hofmann, “Probabilistic Latent Semantic Analysis”, *Proc. Uncertain. Artif. Intell.*, pp. 289-296, 1999, doi: https://doi.org/10.48550/arXiv.1301.6705.

[10] J. C. Campbell, A. Hindle, and E. Stroulia, “Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data”, *Art Sci. Anal. Softw. Data*, Vol. 3, pp. 139–159, 2015, doi: 10.1016/B978-0-12-411519-4.00006-9.

[11] E. Ekinci, “An Aspect-Sentiment Pair Extraction Approach Based on Latent Dirichlet Allocation”, *Int. J. Intell. Syst. Appl. Eng.*, Vol. 3, No. 6, pp. 209–213, 2018, doi: 10.18201/ijisae.2018644779.

[12] K. R. Nastiti, A. F. Hidayatullah, and A. R. Pratama, “Discovering Computer Science Research Topic Trends using Latent Dirichlet Allocation”, *J. Online Inform.*, Vol. 6, No. 1, p. 17, 2021, doi: 10.15575/join.v6i1.636.

[13] S. Suhariyanto, R. Abdullah, R. Sarno, and C. Fatichah, “Aspect-Based Sentiment Analysis for Sentence Types with Implicit Aspect and Explicit Opinion in Restaurant Review Using Grammatical Rules, Hybrid Approach, and SentiCircle”, *Int. J. Intell. Eng. Syst.*, Vol. 14, No. 5, pp. 177–187, 2021, doi: 10.22266/ijies2021.1031.17.

[14] R. Abdullah, S. Suhariyanto, and R. Sarno, “Aspect Based Sentiment Analysis for Explicit and Implicit Aspects in Restaurant Review using Grammatical Rules, Hybrid Approach, and SentiCircle”, *Int. J. Intell. Eng. Syst.*, Vol. 14, No. 5, pp. 294–305, 2021, doi: 10.22266/ijies2021.1031.27.

[15] F. Nurifan, R. Sarno, and K. R. Sungkono, “Aspect Based Sentiment Analysis for Restaurant Reviews Using Hybrid ELMo-Wikipedia and Hybrid Expanded Opinion Lexicon-SentiCircle”, *Int. J. Intell. Eng. Syst.*, Vol. 12, No. 6, pp. 47–58, 2019, doi: 10.22266/ijies2019.1231.05.

[16] A. Berti and W. M. P. V D. Aalst, “A Novel Token-Based Replay Technique to Speed Up

Conformance Checking and Process Enhancement", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol. 12530 LNCS, pp. 1–26, 2021, doi: 10.1007/978-3-662-63079-2_1.

[17] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding", *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, Vol. 1, No. Mlm, pp. 4171–4186, 2019.

[18] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks", In: *Proc. of 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 3982–3992, 2020, doi: 10.18653/v1/d19-1410.

[19] S. Fransiska and A. I. Gufroni, "Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method", *Sci. J. Informatics*, Vol. 7, No. 2, pp. 2407–7658, 2020.

[20] A. A. Farisi, Y. Sibaroni, and S. A. Faraby, "Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier", *J. Phys. Conf. Ser.*, Vol. 1192, No. 1, 2019, doi: 10.1088/1742-6596/1192/1/012024.

[21] J. W. Reed, J. Yu, T. E. Potok, B. A. Klump, M. T. Elmore, and A. R. Hurson, "TF-ICF: A New Term Weighting Scheme for Clustering Dynamic Data Streams", In: *Proc. of 2006 Int. Conf. Mach. Learn. Appl.*, pp. 258–263, 2006, doi: 10.1109/ICMLA.2006.50.

[22] L. H. Suadaa and A. Purwarianti, "Combination of Latent Dirichlet Allocation (LDA) and Term Frequency-Inverse Cluster Frequency (TFxICF) in Indonesian text clustering with labeling", In: *Proc. of 2016 4th International Conference on Information and Communication Technology, ICoICT 2016*, pp. 1–6, 2016, doi: 10.1109/ICoICT.2016.7571885.

[23] S. Yang and H. Zhang, "Text Mining of Twitter Data Using a Latent Dirichlet Allocation Topic Model and Sentiment Analysis Text Mining of Twitter Data Using a Latent Dirichlet Allocation Topic Model and Sentiment Analysis View project", *Int. J. Comput. Inf. Eng.*, Vol. 12, No. 7, pp. 525–529, 2018.

[24] A. Pradhan, M. R. Senapati, and P. K. Sahu, "Improving sentiment analysis with learning concepts from concept, patterns lexicons and negations: Improving sentiment analysis with learning concepts", *Ain Shams Eng. J.*, Vol. 13, No. 2, p. 101559, 2022, doi: 10.1016/j.asej.2021.08.004.

[25] M. Kulmanov, F. Z. Smaili, X. Gao, and R. Hoehndorf, "Semantic similarity and machine learning with ontologies", *Brief. Bioinform.*, Vol. 22, No. 4, pp. 1–18, 2021, doi: 10.1093/bib/bbaa199.

[26] A. Mahmoud and M. Zrigui, "Semantic similarity analysis for paraphrase identification in Arabic texts", In: *Proc. of 31st Pacific Asia Conf. Lang. Inf. Comput.*, No. January, pp. 274–281, 2019.

[27] J. M. Gil, "An overview of textual semantic similarity measures based on web intelligence", *Artif. Intell. Rev.*, Vol. 42, No. 4, pp. 935–943, 2014, doi: 10.1007/s10462-012-9349-8.

[28] M. H. Alam, W. J. Ryu, and S. K. Lee, "Joint multi-grain topic sentiment: modeling semantic aspects for online reviews", *Inf. Sci. (Ny).*, Vol. 339, pp. 206–223, Apr. 2016, doi: 10.1016/J.INS.2016.01.013.