Sentiment Analysis of Hotel Reviews Using Latent Dirichlet Allocation, Semantic Similarity and LSTM

Reza Amalia Priyantina[1]   **Riyanarto Sarno[2]***

[1] *Department of Information Technology Management, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[2,] *Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
* Corresponding author's Email: riyanarto@if.its.ac.id

Abstract: Product reviews are usually determined by sentiment of customers; however sentiment analysis based on aspects still need further research. A hotel commonly has five aspects, which are location, meal, service, comfort and cleanliness. This research proposes methods to determine review sentiment according to the hotel aspects. A hotel reviews are preprocessed into a term list. Firstly, Latent Dirichelet Allocation (LDA) determines the hidden topics of a term list; then Semantic Similarity categorizes the term list based on the topic resulted by Latent Dirichelet Allocation (LDA) into the five aspects of a hotel. Then in calculating similarity measurement, the term list is expanded by using the Term Frequency-Inverse Cluster Frequency (TF-ICF) method. Finally, a classification of customer sentiment (satisfied or dissatisfied) is conducted by using the combination of Word Embedding and Long-short Term Memory (LSTM). The results show that the proposed method can classify the reviews into the five hotel aspects. The highest aspect categorization performance is obtained by using LDA + TF-ICF 100% + Semantic Similarity which reaches 85%; the performance sentiment classification for the highest aspect-based sentiment analysis is obtained by using Word Embedding + LSTM which reaches 93%; and the comfort aspect receives more negative sentiments compared to the sentiments of other aspects. Also the results show that a sentiment is influenced by an aspect.

Keywords: Review analysis, Term list, Aspect categorization, Topic modelling, LDA, LDA-similarity, TF-ICF, Sentiment classification, Word embedding, LSTM.

1. Introduction

The increasing competition in business and the widening range of sales in a company makes the company need to re-examine, compile, organize and improve all aspects related to their business to maintain or improve the companys financial capacity in volatile markets. The ability to increase revenue through customer relationships that are managed effectively is very important for the company to achieve its intended goals [1]. Nowadays, customers can contact companies in easy ways, one of them is through the website on an online review.

Informal communication through internet technology regarding the use and description of certain products or services is avowed as the most influential communication between service providers and consumers, and among consumers [2-5].

Research conducted by [4] found that around 89% of global travelers and 64% of global hoteliers be convinced that online hotel reviews have an effect on hotel bookings. The survey conducted by [6] revealed that almost 95% of tourists made booking decisions by first reading hotel reviews online, and many tourists make online hotel reviews as one of the most important factors in their decision to choose a hotel. Online ratings and customer reviews can be used to help customers make decisions but the reviews give better insight into the hotel [7].

Customer reviews of hotel services and products become a factor for hotel management evaluation and improvement as well as a trigger factor to correct errors and deficiencies in products or services of the hotel [8]. Customer reviews written on the website are large and unstructured data, so that they can be analyzed using appropriate techniques [9].

143

In this study, the researchers conducted an experiment to determine the best classification performance by comparing several review classification techniques. To perform aspect categorization, the researchers used the Latent Dirichelet Allocation (LDA) method, one of the topic modeling methods capable to determine the hidden topic of a document. Such a hidden topic is the important words or the often appearing words in the corpus calculated based on the probability of proximity of topic-document and word-topic [10-12]. In this study, the Latent Dirichlet Allocation (LDA) method was used to determine the hidden topic according to the function of the Latent Dirichlet Allocation (LDA) itself followed by the classification of the predetermined 5 hotel aspects.

Latent Dirichlet Allocation (LDA) has disadvantages in classifying documents into one aspect directly [13]. This study proposed a method which combined the Latent Dirichlet Allocation (LDA) (to find the term list) and the Semantic Similarity to classify a document into the predetermined 5 hotel aspects (location, food, service, comfort, and cleanliness. These hotel aspects were obtained from the Traveloka website on 5 types of rating data and reviews usually complained or appraised by hotel customers. The Semantic Similarity method is a measurement method which defines each document or term which has a distance based on semantic meaning [14, 15]. The expansion of the term list when measuring similarity used the Term Frequency-Inverse Cluster Frequency (TF-ICF) method.

In this study, the experiment to determine the best performance on aspect categorization in was conducted by combining the Latent Dirichlet Allocation (LDA) with semantic similarity using the term list data in Table 1; combining the Latent Dirichlet Allocation (LDA) with semantic similarity using the data term list in Table 1 plus TF-ICF 20%; and combining the Latent Dirichlet Allocation (LDA) with semantic similarity by using data term list in Table 1 plus TF-ICF 100%. The result of the performance of each trial would be discussed further on results and analysis.

After determining the aspect of each review, the sentiment classification would be carried out. The sentiment classification process was carried out in several ways. To conduct sentiment classification, the researchers started by collecting data on aspect categorization based on similar aspect to conduct training data, and obtained the term list sentiment per aspect. The Word Embedding was used to extract each word and document into a word vector used as input for the sentiment classification process using the Long-short Term Memory (LSTM) method. The Long-short Term Memory (LSTM) is a popular method for sentiment classification. The Long-short Term Memory (LSTM) has a "Forget Gate" which functions to decide which information should be stored or discarded and renewed its memory cell [16, 17]. The experiment to determine the best performance in the sentiment classification was carried out by: conducting sentiment classification using Word Embedding + LSTM; sentiment classification using Sentiwordnet; and sentiment classification using LSTM.

This study aims to determine the best performance in each method of analysis and to analyze customer reviews to find out which aspect of the hotel makes customers satisfied and dissatisfied based on the 5 predetermined hotel aspects so that the company was able to plan changes, rearranged its business, and revised their mistakes.

The result of the study was to determine the performance of the best methods to be used in aspect categorization and sentiment classification, to analyze the satisfaction and dissatisfaction with the 5 hotel aspects. The performance result would be evaluated using precision, recall, and F1-Score. It was found that the highest aspect performance category was obtained using the LDA + TF-ICF 100% + Semantic Similarity; the performance of the sentiment classification for the highest aspect-based sentiment analysis was obtained using Word Embedding + LSTM; and the comfort aspect received high negative sentiments compared to other aspects, thus improvements need to be made. Also the results show that a sentiment is influenced by an aspect.

2. Related theory

This section will be explained about several theories related to the research.

2.1 Keyword term list for hotel aspect

In this research, there will be 5 (five) kinds of aspect reviewed, there are cleanliness, comfort, service, food or meal and location which these aspects is taken from the fifth aspect ratings data review on Traveloka application. In addition to the five aspects, the researchers also reviewed on the satisfaction of dissatisfaction of Hotel Customers.

After determining the aspects to be reviewed, the next step is to classify the customer reviews into predetermined aspects. Some keywords that often appear in hotel review texts on every aspects is defined by [7] combined with data from [18] has follows:

Table 1. Keyword aspect term for hotel aspect

Aspect	Keyword Aspect Term for Hotel Aspect
Cleanliness	ventilation, cleanliness, smell, cobweb, smoke, carpet, laundry, furniture, wall, housekeeping, toilet
Comfort	connection, sleep, meeting, charge, activity, bedroom, comfort, feel.
Service	facility, desk, reliable, fast, convenient, wifi, internet check-in, check-out, good, staff, polite, helpful, friendly, reliable, pool, quick, parking, conference room, fee, gym
Food	cafe, drink, breakfast, spicy, meal, bagel, tea, buffet, bar, waffle, restaurant, dinner, lunch, brunch, delicious, food, dish, wine, salad, coffee, pastry, menu, item, cup
Location	location, railway, view, station, airport, distance, far, close, convenient, train, metro, place, mall

Table 2. Pre-processing process

Pre-Processing	Information
Tokenization	At this stage, input text of the document is broken down into the smallest atomic unit. Usually the unit is the form of words or sentences or paragraphs.
Normalization	This process changes all letters into lowercase.
Remove Punctuation	It will be removed punctuation in the sentences.
Stemming	Stemming has a role to make the text into a basic word.
Stopword Removal	Stopwords removal has the role to delete the stop words in English Language.
Spelling Correction	This process has a function to correct any sentences that have errors in writing.

2.2 Pre-processing

One of the natural language processing techniques that are often used in the pre-processing consists of the following techniques, such as:

2.3 Latent Dirichlet allocation (LDA)

Latent Dirichelet Allocation (LDA) [19, 20] is a model that is capable of finding hidden topic of words that appear in the corpus of each document. For each document d on a corpus D, Latent Dirichlet Allocation (LDA) assumes the following processes:

1. Length of Nd sample document from a Poisson Poiss distribution (ξ)

2. Take the $\vec{\theta}_d$ topic distribution from a Dirichlet Distribution Dir (α)
3. For each n word in Nd words:
   a. Select a zd,n topic from the Mult multinomial distribution ($\vec{\theta}_d$)
   b. Select a wd,n word from the Mult multinomial distribution ($\vec{\varphi}_{z_{d,n}}$) with the topic distribution ($\vec{\varphi}_{z_{d,n}}$) from Dir Distribution (β)

According to [21], Latent Dirichlet Allocation (LDA) is defined as

$$p(w, z / \alpha, \beta) = p(w / \alpha, \beta) \, p(z / \alpha). \quad (1)$$

α and β parameters are the model of parameter, w is the word, and z is the topic, $p(z / \alpha)$ is the probability of the topic z found in document d, and $p(w | z, \beta)$ is the probability of the letter w is in topic z. However, computing using Eq. (1) is quite difficult to control. Therefore, Griffith used the approximate equation as follows:

$$p(z_i = k / z_{-i}, w) = \frac{n_{-i,k}^{w_i} + \beta}{n_{-i,k} + W\beta} \cdot \frac{n_{-i,k}^{d_i} + \alpha}{n_{-i}^{d_i} + T\alpha} \quad (2)$$

On Eq. (5), $n_{-i}^{(.)}$ is the total of topic z_i, $n_{-i}^{w_i}$ is the total of topic z that is related to w_i, $n_{-i}^{d_i}$ is the total of topic z that is related to the di documents d_i, W is the number of different pre-processed letters, and T is number of topics. α represents the document-topic density, and β represents topic-word density.

2.4 Semantic similarity

Semantic similarity is a measurement method that defines each document or term which has a distance based on semantic meaning [14, 15]. There are two types of similarity calculation, which are based on the existing resources, such as thesaurus and based on the spread of words in a corpus [22].

$$Similarity(w_i, w_j) = \frac{\sum_{m-1}^{K} w_i^m w_j^m}{\sqrt{\sum_{m-1}^{K} (w_i^m)^2} \sqrt{\sum_{m-1}^{K} (w_j^m)^2}} \quad (3)$$

Similarity distance measures the distance of proximity between word1 (w_i) dan word2 (w_j). $\sum_{m-1}^{K}$ = the number of iterations m to the K word. The maximum value of the similarity distance is 1, which means it is exactly the same, and the minimum value is -1, which means it is completely different. The value of similarity ranged between 0 and 1.

The auxiliary tool used to expand the term and categorize each document is WordNet similarity.

2.5 Expanding term list (TF-ICF)

Term Frequency-Inverse Cluster Frequency (TF-ICF) [23] is one of the weighting terms based on information from documents on a cluster. It makes this method capable of knowing information about the value of a term in a term. In general, Term Frequency-Inverse Cluster Frequency (TF-ICF) perceives the term frequency of the cluster using Eq. (4) where the Inverse Cluster Frequency (ICF) value in term i is influenced by the number of clusters that exists and the number of clusters that contains term i (cfi).

$$ICF_i = 1 + log \frac{c}{cf_i} \qquad (4)$$

While the Term Frequency-Inverse Cluster Frequency (TF-ICF) weight value of each term i in each cluster will be calculated using Eq. (5), where TF_{ji} is the frequency of term i in cluster j and $ICFi$ is the icf value of that term.

$$TF - ICF_j = TF_{ji} \times ICF_i \qquad (5)$$

$$TF - ICF = TF_{t,i} \times \log\left(\frac{N}{CF_t}\right) \qquad (6)$$

Which:

$TF_{t,i}$ = Total terms in class i
N = Total classes
CF_t = Total classes that contains term t

2.6 Word embedding + long-short term memory (LSTM)

In this study, researchers used Word Embedding to extract each word and document into a word vector which will then be used as input for the sentiment classification process using the Long-Short Term Memory (LSTM) method. Word embedding used is GloVe. GloVe is also used to be a tool to help expand the term list during training data and during classification. Glove used as an algorithm to get a global context. So, GloVe has a good accuracy to be applied, because the glove is able to estimate with a good representation of the word [24].

Long-Short Term Memory (LSTM) [25] is recurrent neural network architecture (RNN) designed for modeling the connection between terms with long intervals. The LSTM classification output
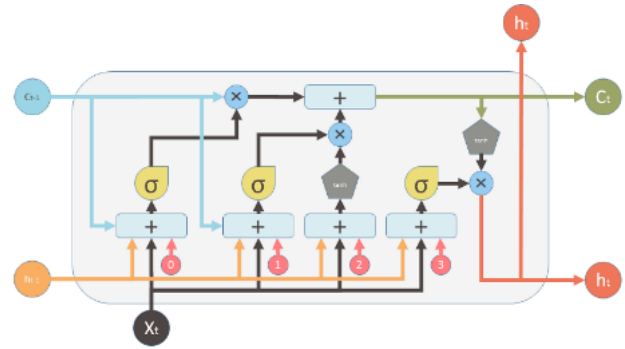

Figure. 1 Long-short term memory architecture

depends on the training data provided. Long-short Term Memory (LSTM) memory can decide whether it would store or discard an information (e.g. Wheter it opens the gate or not), based on interests assigned to the information. Assignment of interests occurs through weights, which were also studied by algorithms. It means that Long-short Term Memory (LSTM) learns from time to time which information is important and which one is not [26].

In general, the architecture of Long-short Term Memory (LSTM) is illustrated in Fig 1. LSTM has been widely used in human language processing such as sentiment analysis, syntactic parsing, long document categorization, etc [16]. In general, the architecture of LSTM is illustrated in Fig 1.

Long-short Term Memory (LSTM) consists of four components namely memory cell (*c*), input gate (*i*) to control the inputs that entering into neurons, output gate (*o*) to control the effect of the activation of neurons, and forget gate (*f*) that makes the neuron is in reset status of its previous status [27]. In general, Long-short Term Memory (LSTM) consists of the following functions:

$$i_t = \sigma(W^{(i)}x_t + U^{(i)}h_{t-1} + b^{(i)}) \qquad (7)$$
$$o_t = \sigma(W^{(O)}x_t + U^{(o)}h_{t-1} + b^{(o)}) \qquad (8)$$
$$f_t = \sigma(W^{(f)}x_t + U^{(f)}h_{t-1} + b^{(f)}) \qquad (9)$$
$$u_t = tanh(W^{(u)}x_t + U^{(u)}h_{t-1} + b^{(u)}) \quad (10)$$
$$c_t = i_t \odot u_t + f_t \odot c_{t-1} \qquad (11)$$
$$h_t = o_t \odot tanh(c_t) \qquad (12)$$

Which:

W, U = matrix weight between two consecutive hidden layers, between the input and hidden layers and between two consecutive cell activation, respectively, that are connected to the gate
⊙ = product value for each element of two vectors
σ = gate sigmoid activation
c = internal state vector
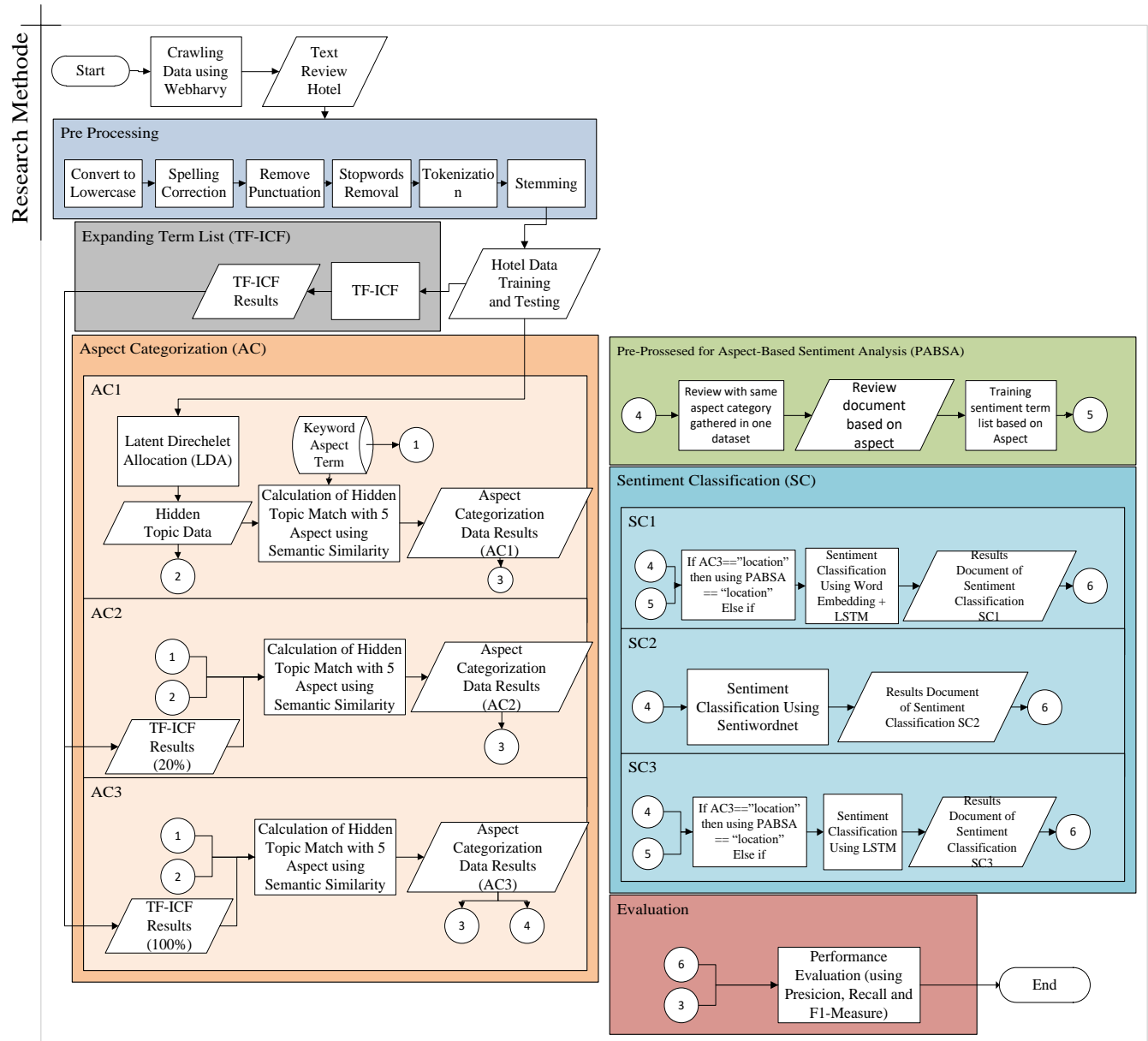h = output cell activation, usually has tanh value

Figure. 2 Research method

2.7 Evaluation

This study proposes a measurement method using Precision, Recall, and F-1 measure. This measurement method is considered the most reliable in measuring the performance of the proposed method [28]. The calculation process of precision and recall is determined by the prediction of information on the actual value represented by True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).

1. Precision is the accuracy level between the information requested by the user and the answer given by the system

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (13)$$

2. Recall is the success rate of the system in rediscovering information

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (14)$$

3. F1 Measure is

$$\text{F1-Measure} = 2 \times \frac{Precision \times Recall}{Precision+Recall} \qquad (15)$$

3. Research method

Firstly, the data were explored using webHarvy on Traveloka. Furthermore, they were pre-processed to clean the terms in review of noise or less relevant text structure for processing. The next module was

Table 3. Representation of the document review in the initial form before the preprocessing stage

IDReview	IDSentence	Text Review
1	1	Spacious guest room with comfortable light and ambiance
2	1	Excellent of breakfast buffet variety

Table 4. Data collection

IDRe view	IDSen tence	Review
1	1	I can hear bass sound
2	2	They like grilled bass
3	3	Very clear sound a bathtub…
4	4	A little fresh air might clear some..
5	5	It was clear that he do not...
6	6	Very awesome and polite made …
7	6	They gave us ideas on tourist..
8	7	Excellent service! Got our room..

the expanding term list (TF-ICF) module for the calculation on semantic similarity at Aspect Categorization. This model aimed to overcome the occurrence of Out-of-Vocabulary, a condition in which the declaration of term list on every aspect is limited, resulting in important terms being undescribed.

Afterwards, the pre-processed data and the expanding term list were included in the aspect categorization module. In this module, the data were classified into 5 aspects: location, food, service, comfort, and cleanliness. The aspect categorization data were processed on the Pre-Processed for Aspect Based Sentiment Analysis. This process collected the aspect categorization result based on the same category. Data of similar aspect would be trained to obtain the term list sentiment per aspect category. The aspect categorization and term list data from the sentiment training were processed to determine the customer satisfaction sentiment classification based on the aspects.

The aspect categorization and sentiment classification were carried out with several experiments (AC1, AC2, AC3 for aspect categorization experiments and SC1, SC2, and SC3 for sentiment classification). The aspect categorization and sentiment classification processes were evaluated to determine the best performance. By using the best performance, it is expected that the end result of the "analysis of customer reviews to determine which hotel aspects make customers satisfied or dissatisfied based on 5 aspects" is well represented.

3.1 Data collection

Firstly, the data were explored using webHarvy on Traveloka.com; for example, "Spacious guest room with comfortable light and ambiance. Excellent variety of breakfast buffet" was processed into:

Based on the given example, it can be seen that the data is a review in the form of sentence, but the reviewer split the data into 2 IDs, meaning that the reviewer separated each review into two point. The ID Review in the Table is the identity of the review obtained, while the Sentence ID is the identity of the sentence split by the researchers.

The example of the review data to be analyzed by the researchers can be seen in Table 4 below.

3.2 Pre-processing

The process taken for this step consists of:

Start

1. Take a review that has been crawled in the previous stage (Table 4).
2. Perform preprocessing using Python Natural Language Toolkit (NLTK).
3. Perform preprocessing in Table 2 with the sequence: Convert into Lowercase → Spelling Correction → Remove Punctuation → Stopwords Removal → Tokenization → Stemming.
4. Save the results of pre-processing.

End

The end result of the pre-processing is in the form of:

Table 5. The results of pre-processing data

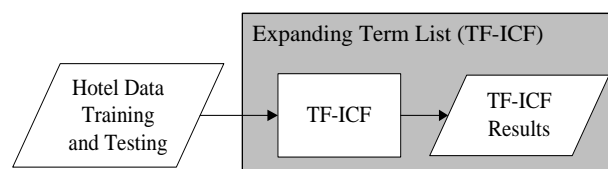| Review	Term List Results
Review 1	'bass', 'hear', 'sound'
Review 2	'bass', 'like', 'grill'
Review 3	'sound', 'bathtub', 'clear', 'easy', 'use', 'very'
Review 4	'clear', 'air', 'fresh', 'little', 'might', 'some'
Review 5	'clear', 'very', 'but', 'disappoint', 'do', 'poor', 'say', 'service', 'understand', 'wish'

Figure. 3 Expanding term list process

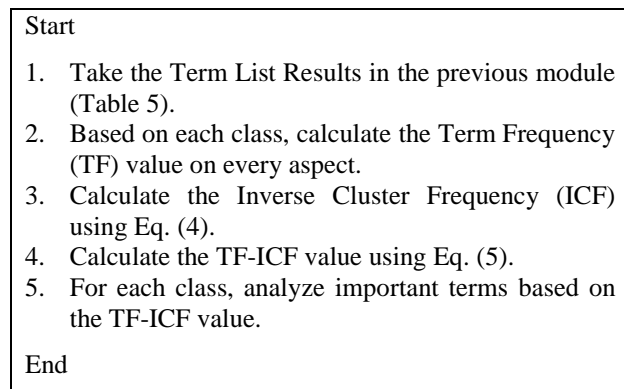## 3.3 Expanding term list (TF-ICF)

This module is intended to overcome the Out-of-Vocabulary problems which might arise during term list declaration. This study extended the term list using the Term Frequency-Inverse Cluster Frequency (TF-ICF). Such a method was used to obtain

Table 6. Calculation of TF-ICF

id	term	TF1	TF2	TF3	TF4	TF5	ICF	TF-ICF1	TF-ICF2	TF-ICF3	TF-ICF4	TF-ICF5
1	excellent	1	0	3	0	0	0.9162	0.9162	0	2.7488	0	0
2	location	1	0	0	1	0	0.9162	0.9162	0	0	0.9162	0
3	hotel	2	0	3	2	0	0.5108	1.0216	0	1.5324	1.0216	0
4	well	1	0	0	0	0	1.6094	1.6094	0	0	0	0
5	locate	3	0	0	0	0	1.6094	4.8283	0	0	0	0
6	look	2	0	2	2	0	0.5108	1.0216	0	1.0216	1.0216	0
7	nice	1	0	0	0	1	0.9162	0.9162	0	0	0	0.9162

important terms for each class. The use of additional term list based on classes which had been marked is expected to improve performance. Thus, it is expected that this module can optimize the "categorization process" module, especially when the hidden topic calculation was matched with aspects of the semantic similarity process.

There are steps from the Term Frequency-Inverse Cluster Frequency (TF-ICF) to determine additional term list, explained in the following pseudocode:

```
Start

1.  Take the Term List Results in the previous module
    (Table 5).
2.  Based on each class, calculate the Term Frequency
    (TF) value on every aspect.
3.  Calculate the Inverse Cluster Frequency (ICF)
    using Eq. (4).
4.  Calculate the TF-ICF value using Eq. (5).
5.  For each class, analyze important terms based on
    the TF-ICF value.

End
```

Here is the result of the calculation of the expanding term list:

TF1 shows the number of term occurrences in cluster 1, TF2 in cluster 2, and so on until TF5 which shows the number of term occurrences in the cluster 5. ICF considers the occurrence of each term in the cluster based on logarithmic function. If the ICF value is 0, it means that the term continues to appear in all clusters. When the logarithmic function calculates the result of the division of the number of clusters with the cluster of occurrences of terms, it will produce a log (1) value, which is zero. ICF shows that the fewer clusters where the term appears, the ICF value will tend to be large, as in ID 4 and 5.

The important terms of each cluster were determined by calculating the TF-ICF value. Table 5 TF-ICF1 shows the value of a term in cluster 1; TF-
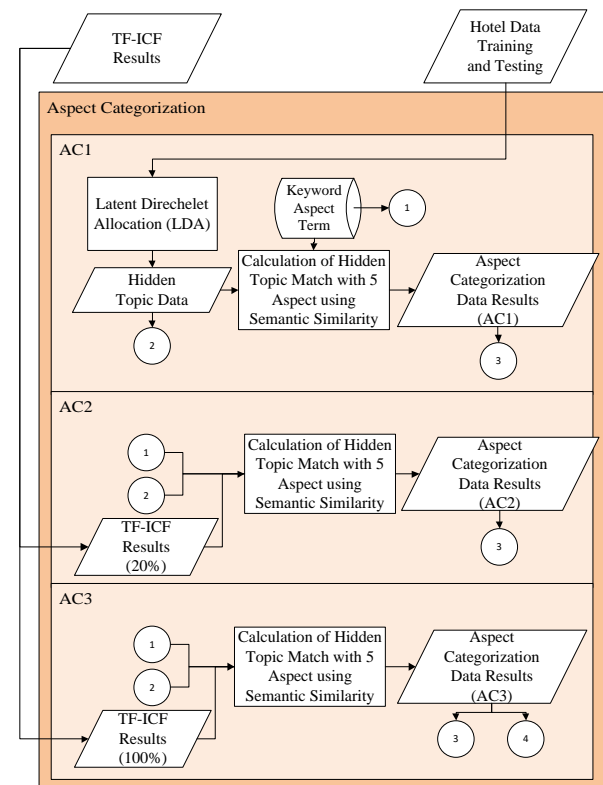

Figure. 4 Aspect categorization performance analysis

ICF2 shows the value of a term in cluster 2, and so on until TF-ICF5 shows the value of a term in cluster 5. The Table shows that the frequency value of a term in each cluster is able to affect the ICF value in producing the final value of TF-ICF; the higher the frequency value of a term, as long as the ICF value is not zero, the TF-ICF value generated is greater (for example, ID 4 and 5 data).

3.4 Aspect categorization

The researchers conducted several experiments to determine the best aspect categorization performance. Every experiment carried out hereinafter will be called Aspect Categorization (AC). AC1 used the Latent Dirichlet Allocation (LDA) method to

produce hidden topic data. The hidden topic data were calculated using the Semantic Similarity method to categorize them into the 5 predetermined hotel aspects in Table 1; AC2 = AC1 + TF-ICF 20% and AC3 = AC1 + TF-ICF 100%.

The description of the Aspect Categorization experiment can be seen as follows:

3.4.1. Aspect categorization (AC) 1

This experiment was carried out using the Latent Dirichlet Allocation (LDA) method to search for the hidden topics followed by semantic similarity using data in Table 1 to categorize each review. The tool used to expand the term and categorize each document was the WordNet similarity. The description of AC1 can be seen below:

Latent Dirichlet Allocation (LDA)

The clustering method used was Latent Dirichlet Allocation (LDA). The result obtained from this process was in the form of hidden topics and the probability values from the input documents.

The hidden topics obtained were the result of topic identification from words, documents, and the whole corpus. The number of hidden topics issued was not predetermined due to the varied length of documents. For each Latent Dirichlet Allocation (LDA) process, the researchers conducted:

Start

1. Take term list that have been pre-processed in the previous module (Table 5).
2. By using Python Gensim, calculate the probability of appearance of each word in the document and corpus and also the probability of proximity of document-topic and word-topic (using Eq. (2)).
3. Show hidden topic data (important words or words that appear a lot in a corpus).

End

The result of hidden topic calculation using Latent Dirichlet Allocation (LDA) from each term list document is presented as follows:

Table 7. LDA Hidden Topic Results

Term List	Hidden Topics
Term List 1	['0.070 × sound' + '0.050 × hear' + '0.070 × bass']
Term List 2	['0.059 × like' + '0.079 × bass']
Term List 3	['0.070 × sound' + '0.068 × bathtub' + '0.081 × clear' + '0.083 × easy' + '0.055 × use']
Term List 4	['0.019 × clear' + '0.060 × air' + '0.089 × fresh' + '0.019 × little']
Term List n	Hidden Topics (n)

After obtaining the hidden topic for each document, the result along with the term list in Table 1 was processed in terms of its similarity calculation using the Semantic Similarity. This calculation was conducted to categorize each review on the 5 predetermined hotel aspects.

Aspect Categorization using Semantic Similarity

The Hidden Topic Similarity process is the last step in the Aspect Categorization process. The input is the hidden topic described in Table 7 which had been identified using the Latent Dirichlet Allocation (LDA) method followed by similarity calculation using the data term list in Table 1 to classify the review of the predetermined 5 hotel aspects.

The highest similarity value was used as the class label of the document. The scale of the value ranged between 1 and 0. If the value is closer to 1, the hidden topic is really similar and most likely able to represent the aspect. Otherwise, the value 0 represents that the hidden topic is not similar and cannot represent the aspect. The similarity which would be analyzed consisted of 5 aspect classifications: location, food, service, comfort, and cleanliness. Using the Semantic Similarity formula on Eq. (3), the result in Table 8 below is obtained from the calculation of the LDA + Semantic Similarity analysis:

Table 8. LDA + semantic similarity analysis result

Term List 1	Aspect 1 Location	Aspect 2 Meal	Aspect 3 Service	Aspect 4 Comfort	Aspect 5 Cleaness
Hidden Topic : ['Sound', 'Hear', 'Bass']. Aspect should be: Comfort					
Sound	0.038	0.000	0.041	0.047	0.000
Hear	0.000	0.000	0.041	0.047	0.000
Bass	0.038	0.087	0.000	0.000	0.084
Total	0.075	0.087	0.082	**0.094**	0.084
Term List 2	Aspect 1 Location	Aspect 2 Meal	Aspect 3 Service	Aspect 4 Comfort	Aspect 5 Cleaness
Hidden Topic : ['Like', 'Bass']. Aspect should be: Meal					
Like	0.018	0.000	0.019	0.022	0.039
Bass	0.018	0.050	0.019	0.022	0.000
Total	0.035	**0.050**	0.038	0.045	0.039
Term List 3	Aspect 1 Location	Aspect 2 Meal	Aspect 3 Service	Aspect 4 Comfort	Aspect 5 Cleaness
Hidden Topic : ['Sound', 'Bathtub', 'Clear', 'Easy', 'Use']. Aspect should be: Comfort					
Sound	0.000	0.000	0.022	0.026	0.032
Bathu	0.000	0.000	0.022	0.026	0.000
Clear	0.029	0.030	0.000	0.026	0.032
Easy	0.029	0.000	0.022	0.000	0.000
Use	0.000	0.030	0.000	0.000	0.000
Total	0.058	0.060	0.067	**0.079**	0.064

From Table 8, it can be seen that, for Term List 1, the number closest to 1 is in Aspect 4 (marked yellow in the total column), "Comfort". It can be seen that based on the data in Review 1 in Table 5 and in Table 6, the annotators agreed to give aspect to the Review 1 data with the "Comfort" aspect. In Term List 2, the biggest number on Aspect 2 or meal can be checked in the same way and the annotators also agreed to give the Review 2 as the "meal" aspect. As well as Term List 3, the "Comfort" aspect, the same label as the one given by the annotator in Review 3.

3.4.2. Aspect Categorization (AC) 2

Aspect Categorization (AC) 2 underwent the same process as AC1. However, the term list used to perform the Semantic Similarity in the AC2 did not only use the term list in Table 1 but also the additional data from the Expanded Term list of TF-ICF 20%.

3.4.3. Aspect Categorization (AC) 3

Aspect Categorization (AC) 2 underwent the same process as AC1 and AC2. However, the term list used to perform the Semantic Similarity in the AC2 did not only use the term list in Table 1 but also the additional data from the Expanded Term list of TF-ICF 100%.

3.5 Pre-prossesed for aspect based sentiment analysis

At this stage, the data from the Aspect Categorization (Table 9) were categorized based on the same aspect categories (Table 10 and Table 11). Afterwards, training was conducted to obtain the term list sentiment from every aspect by labeling each data with the desired sentiment for the data. Training data aimed to train review data on each aspect of positive or negative sentiment. Sentiment labeling for training was carried out by the annotator.
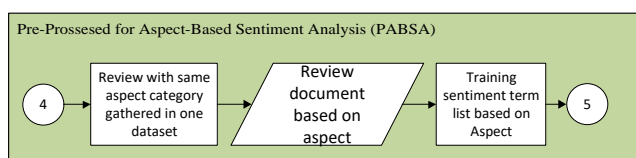

Figure. 5 Pre-prossesed for aspect based sentiment analysis

Table 9. Aspect Categorization Data Review

No.	Term List	Aspect
1.	'Sound', 'Hear', 'Bass'	Comfort
2.	"Like", "bass"	Meal
3.	"Sound", "Bathtub", "Clear", "Easy", "Use"	Comfort

Table 10. Training Data for Aspect-Based Sentiment (Comfort Aspect)

No.	Term List	Aspect	Sentiment
1.	'Sound', 'Hear', 'Bass'	Comfort	Negative
2.	"Sound", "Bathtub", "Clear", "Easy", "Use"	Comfort	Positive

Table 11. Training Data for Aspect-Based Sentiment (Meal Aspect)

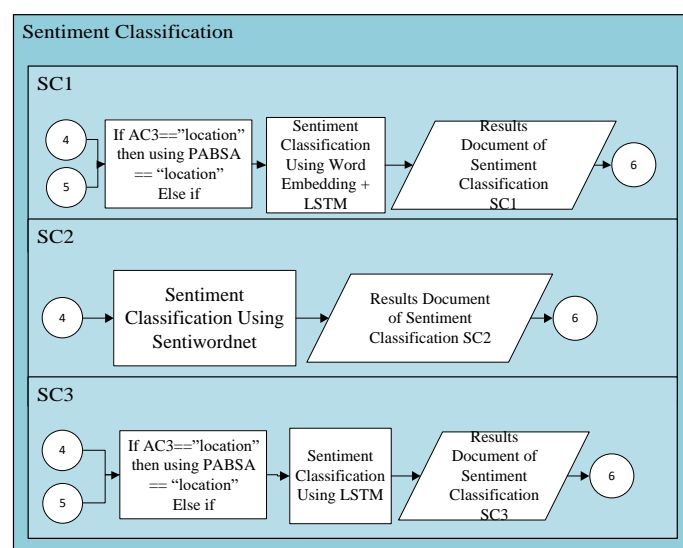| No.	Term List	Aspect	Sentiment
1.	"Like", "bass"	Meal	Positive

Figure. 6 Sentiment classification using word embedding + LSTM

3.6 Sentiment classification

In conducting an experiment to determine the best Sentiment Classification (SC) performance, SC1 was conducted a prior inspection on aspect; if, for example the aspect which follows the data is "location", the review will be classified using the term list sentiment data from "location" training. The classification process used Word Embedding + LSTM; SC2 = SC1, but the classification used Sentiwordnet; and SC3 = SC1 classification used LSTM.

The description of the Sentiment Classification experiment can be seen as follows:

3.6.1. Sentiment classification (SC) 1

The following processes were conducted for the sentiment analysis:

Start

1. Hotel aspect categorization data obtained using the LDA + Semantic Similarity;

2. Determining aspect following the review data;
3. Classifying the sentiment based on the sentiment data training of the same aspect;
4. Researchers conducting Word Embedding to extract each word into vector followed by term list expansion;
5. Classifying the sentiment using the LSTM formula from Eq. (7) to Eq. (12);
6. Presenting the positive and negative sentiment analysis result of each term list.
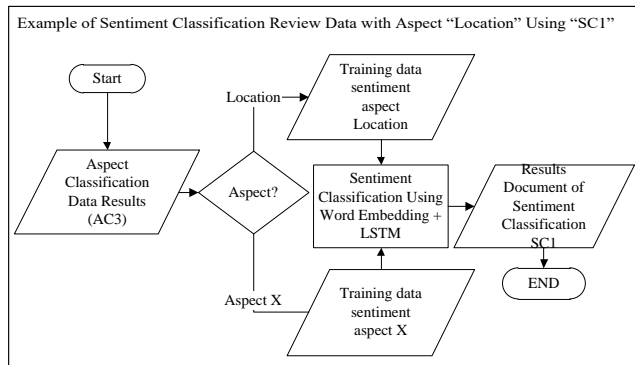
End

Figure. 7 Example of sentiment classification review data with "Location" aspect using "SC1"

In conducting Sentiment Classification, the researchers underwent the following detail processes shown in Fig. 7.

The example of the data to be processed for aspect-based sentiment classification can be seen in Table 9. The data in Table 9 would be extracted to obtain word vector and to be expanded in terms of the term list using Word Embedding. The Word Embedding used was GloVe. After obtaining the number vector (word vector), the data were classified using Long-short Term Memory (LSTM).

The result of the Sentiment classification using Word Embedding + LSTM can be seen in Table 12 below.

The result of the sentiment classification using Word Embedding + LSTM is in the form of sentiment scores with positive or negative numbers as in Table 12.

Table 12. The Results of Sentiment Classification using Word Embedding + LSTM

| No. | AC3 Data | Aspect | Score Sentiment | |
			Positive	Negative
1	**Term List x**	Comfort	0.345	0.757
2	**Term List y**	Comfort	0.875	0.223
3	**Term List z**	Meal	0.984	0.023

3.6.2. Sentiment Classification (SC) 2

SC2 classified sentiments using SentiwordNet. In this study, the researchers attempted to compare two general methods usually used for sentiment evaluation, one of which is Long-short Term Memory (LSTM) and the other is Sentiwordnet. According to [29] SentiWorNet is one of the important resources in conducting sentiment classification. Sentiwordnet is a lexical resource for extracting opinions; there are 3 subjective weighting for each word: positive, negative and objective [30-31]. In this study, the Sentiwordnet used was the 3.0 version of Sentiwordnet, in which this was an enhanced version and was increased by more than 19% of the old version [32]. The Sentiwordnet method is also considered to be a good method for analyzing sentiment.

Sentiwordnet has polarity which prevents "over-fitting", which occurs when word weighting must be conducted in a particular set of documents and leads to poor performance in the dataset test. The synset polarity can be calculated as follows:

$$Pt,_{swn} = SS_{pos} - SS_{neg} \qquad (16)$$

Which:
SS = Scores from synet are defined in Sentiwordnet
The result of SC2 is also in the form of positive or negative score as in Table 12.

3.6.3. Sentiment Classification (SC) 3

The process of SC3 is similar to SC1. However, the sentiment classification process used Long-short Term Memory (LSTM).

3.7 Evaluation

Evaluation was conducted by comparing a number of performances for Aspect Categorization and Sentiment Classification. Each performance would be evaluated using Precision, Recall, and F1 Score.

The formula of each method can be seen in Eqs. (13), (14) and (15) in this section.

Every performance from Aspect Classification and Sentiment Analysis was compared using the scenario described and would be shortened in the Results and Analysis.

4. Result and analysis

Results and analysis section, the results of the Aspect and Sentiment analysis will be explained.

Table 13. Approach for aspect categorization

No. Approach	Description
AC1	LDA produced term list. The term list was categorized into 5 hotel aspects using Semantic Similarity. Semantic similarity will calculate the similarity of the term list to the aspects written in Table 1.
AC2	AC1 + Expanding term list using TF-ICF, 20% obtained from the total synonym.
AC3	AC1 + Expanding term list using TF-ICF, 100% obtained from the total synonym.

Table 14. Aspect categorization performance

ASPECT ANALYSIS PERFORMANCE		
No. Approach	Performance Metrics	F1-Measure
AC1	LDA + Similarity data on Table 1	0.609
AC2	LDA + TF-ICF (20%) + Similarity	0.656
AC3	LDA + TF-ICF (100%) + Similarity	0.849

Table 15. Approach for sentiment classification

No. Approach	Description
SC1	The approach used data from AC3 aspect classification. Researchers conduct training and testing reviews based on aspects. This training aims to train sentiment classification data on aspect. The data will be classified using Word Embedding + LSTM. The classification results are hotel sentiment based on aspect.
SC2	SC1, but the sentiment classification process used SentiWordNet
SC3	SC1, but the sentiment classification process used LSTM

4.1 Approach for aspect categorization

Based on 529 customer review data crawled from the Traveloka website for The Manhattan at Times Square Hotel, an experiment was conducted to calculate the performance of the method to determine the hidden topic and evaluate the performance for aspect classification using 3 Approach Aspect Categorization testing.

Each Categorization Approach was evaluated using precision, recall, and F1-Measure, each formula can be seen in Eq. (13), Eq. (14) and Eq. (15). The evaluation based on the Aspect Categorization using Latent Dirichlet Allocation (LDA), Term Frequency-Inverse Cluster Frequency (TF-ICF), and Semantic Similarity is as shown in Table 14.

Table 16. Sentiment classification performance

Sentiment Analysis Performance			
Performance Metrics	Score Analysis Performance, Using:		
	SC1	SC2	SC3
Precision	0.906	0.759	0.800
Recall	0.960	0.777	0.988
F1-Measure	0.932	0.767	0.884

From Table 14, it can be seen that the Approach Aspect Categorization 3 (AC3) conducted evaluation using Latent Dirichlet Allocation (LDA) to determine hidden topics then the found hidden topic was calculated in terms of its similarity using Semantic Similarity with the keyword term in Table 1, and the term obtained from Term Frequency-Inverse Cluster Frequency (TF-ICF) 100% calculation to classify aspects of each review given. The Approach Aspect Classification 3 is the best approach which can produce application performance up to 0.849 or 85% compared to the others.

4.2 Approach for sentiment classification

The review data were labeled sentiment by the annotator. The sentiments were divided into negative and positive. If there was an adjective within a neutral sentiment, the annotator would classify the sentiment to be negative or positive based on the rating score. The review data were divided into 80% for training and 20% for testing.

After aspect categorization, the researchers obtained review data along with aspects of each review. Review data and aspects (AC3 result) were analyzed in order to determine the best sentiment classification performance. The performance of the sentiment classification was also carried out with several experimental approaches described as shown in Table 15.

The performance of each sentiment classification carried out was calculated using the same evaluation method as the evaluation method used to evaluate aspect categorization, namely using precision, recall, and F1-Measure, each of which can be seen in Eqs. (13), (14) and (15).

The result of the performance evaluation of each approach can be seen in the following Table 16. From Table 16, it can be seen that aspect-based sentiment analysis using AC3 data with aspect-based sentiment analysis method using Word Embedding + LSTM had a high level of performance of 0.927 or 93%.

Since the Sentiment Classification in SC1 is the

Table 17. The proposed approach

Aspect Based Sentiment Analysis	
AC3	SA1
Latent Dirichlet Allocation (LDA) produced Hidden Topic into a term list. The term list was classified into 5 hotel aspects using Semantic Similarity. Semantic Similarity calculated the similarity between term list and the aspects in Table 1 + Expanding Term List obtained from Term Frequency-Inverse Cluster Frequency (TF-ICF) process. The TF-ICF used was 100% from the total synonym.	Data from AC3 of aspect categorization were used. The researchers conducted data training and testing for sentiment classification of each aspect-based review. This training aimed to train sentiment classification data on the aspects. Furthermore, the data were classified using Word Embedding + LSTM. The classification result was in the form of Aspect-Based Sentiment Analysis Hotel.

Table 18. Sentiment evaluation results on hotels aspect analysis

Sentiment Evaluation Results on Aspect		
Aspect	Sentiment	Evaluation Results (in Percent)
Location	Positive	21.114
	Negative	0.464
Meal	Positive	2.784
	Negative	0.696
Service	Positive	45.940
	Negative	3.016
Comfort	Positive	10.905
	Negative	11.369
Cleanness	Positive	2.552
	Negative	1.160
Total Percentace		100

Table 19. An example of a sentiment change influenced by an aspect

Reviews Example	Aspect	Sentiment
We Can Hear the **Bass** Sound	Comfort	Negative
She Like Grilled **Bass**	Meal	Positive

best method, the researchers suggested conducting the Approach for Sentiment Classification method in SC1 to carry out aspect-based sentiment analysis. Table 17 shows the method to be used to conduct aspect-based sentiment analysis.

4.3 The results of aspect based sentiment analysis hotel evaluation

The following is the final evaluation results of aspect based sentiment analysis for Hotel Management optimization based on customer online reviews using Latent Dirichlet Allocation (LDA) Similarity and Word Embedding LSTM methods.

From Table 18, it can be concluded that hotel customers often provide reviews on the hotel service. The results of hotel service analysis have 45.940 positive reviews and 3.016 negative reviews, which means that the hotel service is very satisfying or need a little improvement to eliminate the 3.016 negative reviews on the hotel service. There are also 21.114 positive reviews and only 0.464 negative reviews on the aspect of location. Slightly different from the aspects of service and location, the aspects of cleanness and meal have slight differences in positive and negative reviews, the positive review on cleanness is 2.552 and the negative review is 1.160. Meanwhile, the aspect of meal has a positive review of 2.784 and a negative review of 0.696. The aspect of comfort is different from the overall review that has been analyzed. On this aspect, the number of negative reviews is greater than the positive ones, where the positive review is 10.905 and the negative review is 11.369.

4.4 Effect of aspect to the sentiment

Aspect-based Sentiment Analysis conducted by the researchers had been concluded in Table 18. From the result, it was found that "the classification of topic into hotel aspects affects the sentiment result." The effect of the aspects on sentiment analysis can be seen in the example as shown in Table 19.

From Table 19, it can be seen that the word "Bass" has 2 different meanings. Bass in the first review means musical instrument and the second one means food. Both "bass", then, have different aspects. The first bass is a comfort aspect because it is a musical instrument, so that it refers to the convenience of the customer, while the second one refers to food aspect based on the sentence.

Even though the sentences used the same word, the sentiment possessed by each review may differ depending on the aspect and sentence given. In the first aspect, "bass" is classified as comfort, so that it is considered as a nuisance because "customers can hear bass sounds", which the customer might feel uncomfortable with because customers usually need comfort and tranquility in the hotel, not loud music or disturbance, thus this review has negative aspect. Meanwhile, the second "Bass" is in the aspect of food; when the customer stated that they liked bass, it means they praised for the food provided by the hotel resulting in positive sentiment.

5. Conclusion

The results show that the proposed method can classify the reviews into the five hotel aspects. The highest aspect categorization performance is obtained by using LDA + TF-ICF 100% + Semantic Similarity which reaches 85%; the performance sentiment classification for the highest aspect-based sentiment analysis is obtained by using Word Embedding + LSTM which reaches 93%; and the results of sentiment evaluation analysis on the topic analysis found that the hotel has a positive sentiment on the aspect of service of 45.940 with negative sentiment of only 3.016, which means customers are satisfied with the service provided by the hotel, the lowest negative sentiment is on the aspect of location, which is 0.464. It means that the hotel is in the right location. The thing to note is that the hotel has a higher negative sentiment than the positive sentiment on the aspect of comfort by 11.369, which means that the hotel management must pay more attention to the comfort for their customer. Also the results show that a sentiment is influenced by an aspect.

Acknowledgments

References

[1] A. Ghazian, M. Hossaini, and H. Farsijani, "The effect of customer relationship management and its significant relationship by customers' reactions in LG Company", In: *Proc. of International Conf. on Economics and Finance*, pp.42 – 50, 2016.

[2] A. Cantallops, and F. Salvi, "New Consumer Behavior: A Review of Research on eWOM and Hotels", *International Journal of Hospitality Management,* Vol.36, No.7, pp.41-51, 2014.

[3] M. Schuckert, R. Law, and L. Xianwei, "Hospitality and Tourism Online Reviews: Recent Trends and Future Directions", *Journal of Travel & Tourism Marketing,* Vol.32, No.5, pp.608-621, 2015.

[4] Y. Hu and K. Chen, "Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings", *International Journal of Information management,* Vol.36, No.3, pp.929-944, 2016.

[5] E. Park, J. Kang, D. Choi, and J. Han, "Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews", *Current Issues in Tourism,* pp.1-7, 2018.

[6] M. Ady and D. Quadri-Felitti, "Consumer Research Identifies Which Attributes Are Most Important to Travelers When Booking A Hotel", Trust You, London, 2015.

[7] N. Akhtar, N. Zubair, A. Kumar, and T. Ahmad, "Aspect based Sentiment Oriented Summarization of Hotel Reviews", In: *Proc. Computer Science*, Vol.115, pp.563-571, 2017.

[8] M. Geetha, P. Singha, and S. Sinha, "Relationship between customer sentiment and online customer ratings for hotels - An empirical analysis", *Tourism Management*, Vol.61, No.22, pp.43-54, 2017.

[9] A. Gandomi, and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics", *International Journal of Information Management*, Vol.35, No.2, pp.137-144, 2015.

[10] Y. Fu, M. Yan, X. Zhang, L. Xu, D. Yang, and Jeffr, "Automated classification of software change messages by semi-supervised Latent Dirichlet Allocation", *Information and Software Technology*, Vol.57, No.17, pp.369-377, 2015.

[11] Y. Heng, Z. Gao, Y. Jiang, and X. Chen, "Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach", *Journal of Retailing and Consumer Services*, Vol.42, No.6, pp.161-168, 2018.

[12] Q. Chen, L. Yao, and J. Yang, "Short text classification based on LDA topic model", *Language and Image Processing,* pp.749-753, 2017.

[13] T. Miller, D. Dligach, and G. Savova, "Unsupervised Document Classification with Informed Topic Models", In: *Proc. of International Conf. of the 15th Workshop on Biomedical Natural Language Processing,* pp.83–91, 2016.

[14] E. Pamungkas, R. Sarno, and A. Munif, "B-BabelNet: Business-Specific Lexical Database for Improving Semantic Analysis of Business Process Models", *Telecommunication, Computing, Electronics and Control*, Vol.15, No.1, pp.407-414, 2017.

[15] C. Tsai, W. Lin, Y. Hu, and G. Yao, "Under-sampling class imbalanced datasets by combining clustering analysis and instance selection", *Information Sciences*, Vol.477, No.29, pp.47-54, 2019.

[16] W. Huang, G. Rao, Z. Feng, and Q. Cong, "LSTM with sentence representations for

Document-level Sentiment Classification", *Neurocomputing*, Vol.308, No.45, pp.49-57, 2018.

[17] H. Do, P. Prasad, A. Maag, and A. Alsadoon, "Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review", *Expert Systems With Applications*, Vol.118, No.3, pp.272-299, 2019.

[18] G. Chen and L. Chen, "Augmenting service recommender systems by incorporating contextual opinions from user reviews", *User Modeling and User-Adapted Interaction*, Vol.24, No.3, pp.1-38, 2015.

[19] A. Baskara, R. Sarno, and A. Solichah, "Discovering traceability between business process and software component using Latent Dirichlet Allocation", *2016 International Conference on Informatics and Computing*, pp.251-256, 2016.

[20] B. Rintyarna, R. Sarno, and A. L. Yuananda, "Automatic ranking system of university based on technology readiness level using LDA-Adaboost.MH", In: *Proc. of International Conf. On Information and Communications Technology*, pp.495-499, 2018.

[21] Y. Fu, M. Yan, X. Zhang, L. Xu, D. Yang, and J. Kymer, "Automated classification of software change messages by semi-supervised Latent Dirichlet Allocation", *Information and Software Technology*, Vol.57, N0.17, pp. 369-377, 2015.

[22] E. Iosif and A. Potamianos, "Similarity computation using semantic networks created from web-harvested data", *Journals Cambridge: Natural Language Engineering,* Vol.21, No.1, pp.49-79, 2013.

[23] L. 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 Conference on Information and Communication Technology*, pp.1-6, 2016.

[24] S. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi, "Sentiment analysis based on improved pre-trained word embeddings", *Expert Systems with Applications,* Vol.117, No.44, pp.139-147, 2019.

[25] M. Heikal, M. Torki, and N. El-Makky, "Sentiment Analysis of Arabic Tweets using Deep Learning", In: *Proc. of International Conf. of procedia Computer Science*, pp.114-122, 2018.

[26] J. Wang, B. Peng, and X. Zhang, "Using a Stacked Residual LSTM Model for Sentiment Intensity Prediction", *Neurocomputing*, Vol.322, No.49, pp.93-101, 2018.

[27] F. Guzman, S. Joty, L. Marquez, and P. Nakov, "Machine translation evaluation with neural networks", *Computer Speech & Language,* Vol.45, No.5, pp.180-200, 2017

[28] Y. Jiao and P. Du, "Performance measures in evaluating machine learning based bioinformatics predictors for classifications", *Higher Education Press and Springer-Verlag Berlin Heidelberg*, Vol.4, No.4, pp.320-330, 2016.

[29] A. Hamouda and M. Rohaim, "Reviews Classification Using Sentiwordnet Lexicon", *The Online Journal on Computer Science and Information Technology,*Vol.2, No.1, pp.120-123, 2011.

[30] M. Fikri and R. Sarno, "A Comparative Study of Sentiment Analysis using SVM and SentiWordNet", *Indonesian Journal of Electrical Engineering and Computer Science*, Vol.13, No.3, pp.902-909, 2019.

[31] S. Suhariyanto, A. Firmanto, and R. Sarno, "Prediction of Movie Sentiment Based on Reviews and Score on Rotten Tomatoes Using SentiWordnet", In: *Proc. of 2018 International Seminar on Application for Technology of Information and Communication*, pp.202-206, 2018.

[32] M. AL-Sharuee, F. Liu, and M. Pratama, "Sentiment analysis: An automatic contextual analysis and ensemble clustering approach and comparison", *Data and Knowledge Engineering*, Vol.115, No.1, pp.194-213, 2018.