

THE SENTIMENT CLASSIFICATION OF HOTEL REVIEWS AND HOTEL DESCRIPTION USING FEATURE-BASED TECHNIQUE FOR CUSTOMER RELATIONSHIP MANAGEMENT

MONTAKARN TUMMAWAT AND CHATKLAW JAREANPON*

Polar Lab
Department of Computer Science
Faculty of Informatics
Mahasarakham University
Khamriang Sub-District, Kantharawichai, Mahasarakham 44150, Thailand
58011260502@msu.ac.th; *Corresponding author: chatklaw.j@msu.ac.th

Received June 2022; accepted August 2022

ABSTRACT. *Learning and understanding customer needs is one of the business strategies that will help build long-term customer relationships. This research has analyzed customer opinions compared to hotel features, and allowed the hoteliers to use this information to develop and improve their business to meet the needs of their guests. This research proposed: 1) compilation of English comments from the website, 2) word segmentation process consists of labeling the types of words using the Penn Treebank Target and extracting the types of words that are important to the analysis as follows: verbs, adjectives, and adverbs to be processed, 3) the customer feedback analysis process is used to identify the feedback poles of each feature, 4) extracting the hotel description, and 5) feature matching between hotel description and prediction result. It uses to check the consistency between the customer reviews and hotel strengths. The results showed that the efficacy of the analysis of hotel guest reviews with the highest and average F-measure values were 0.83 and 0.56, respectively.*

Keywords: Opinion mining, Sentiment analysis, Customer relationship management, Text mining

1. Introduction. Customer Relationship Management (CRM) is a business strategy to build long-term relationships with customers. Learning the different needs of products or services of customers is suitable for creating customer satisfaction. An efficient management system has been developed by customer satisfaction [1]. At present, customer relationship management also uses new information technology devices to interact [2] called electronic Customer Relationship Management (eCRM) to provide the users with impressive service. It aims to provide a positive experience for the customer and builds the loyalty of products and services called Customer Experience Management (CEM). Nair et al. [3] developed an effective method to tackle the hotel management system for maintaining records of the guests requirement and create the automated system for all hotel services and email-based communication system to retain the customers. [4] proposed the techniques for analyzing various customer retention to enhance customer relationships through data mining, and it helped and understood for buying pattern and classified customers. The eCRM is a meeting place to communicate and comment on overall, for example in hotel, such as “This hotel is good”, “This hotel is okay”, and “This hotel is bad”. There are also reviews or mentions of the different expectations of the hotel, such as “Clean rooms”, and “Easy to get around”. These comments generate a lot of information and attitudes of hotel users. The hotel can use the information and analysis to improve and develop the hotel efficiency. It is also an essential resource for

tourists. Therefore, many hotel property factors must be considered, such as rooms, services, location, food, and cleanliness, by analyzing reviews from those who have stayed or used the service. It will enable the development of a CRM system to identify guests who are interested in hotel properties [5], offer a new way to classify the opinions at the sentence level, and categorize the opinions divided into positive or negative reviews [6]. They provided a method for analyzing customer reviews about hotels to examine factors affecting customer hotel selection [7], analyzing and predicting the satisfaction scores from customer reviews [8]. Ennaji et al. [9] presented a social intelligence framework to help entrepreneurs know more about customer opinions on products. It also helped the organizations to forecast their market share. Afzaal and Usman [10] presented a framework that extracted tourism data from Twitter, analyzed the data from various perspectives, and visualized the results of the analysis as positive and negative sentiments. [11] presented a value concept model that described the theoretical link between the CRM dimensions and hotel performance. That can explain the mediating role of marketability in the relationship. Bucur [12] presented an online tourism-related platform to collect and summarize hotel service reviews to inform customers who are looking for accommodation and enable hotels to use the information to improve their services. The proposed system extracted hotel reviews from the Internet and categorized them using opinion mining techniques. The algorithm's performance was analyzed using text-mining domain-specific measures and proposed methods to improve results. Agarwal et al. [13] presented a new sentiment analysis model extracted from ConceptNet based ontology and context information to define domain-specific concepts that generate domain-specific key attributes. Additionally, the polarity of isolated concepts was determined using contextual polar dictionaries. This work has been developed by considering the context information of words. Iriani et al. [14] proposed a method using deep learning to classify the utterances of hate, extremism, and radicalism which are posted on social media.

The previous research found that most research analyzes the customer opinions about a product or service to extract their opinions. However, it does not appear to interest in the hotel description that it is able to controversial content, for example, the hotel description is "nice location" but the opinion of the location is negative. This conflict is affected by the hotel's image. Therefore, in this research, a method of analyzing hotel guest reviews and hotel descriptions was proposed and presented. The hotel business knows the needs of hotel guests and brings the positive and negative comments for improving the ineffective management, meeting guests' needs, and trying to increase the market share for the hotel business.

2. Problem Statement and Preliminaries. From the various research studies [3-7], they proposed the process of developing opinion classification and analyzing property data of the commenters such as gender, age, occupation, income, and education. Businesses and organizations develop it to help and meet the needs of service users including the customer relationship management, a part of the CRM itself. The data sources used for analysis from social networking sites include Facebook pages, Twitter blogs, hotel websites such as Agoda (www.agoda.com), TripAdvisor (www.tripadvisor.com), and venere (www.venere.com) UCI databases. The tools to collect data from Facebook are Graph API. The PageFaceger Import.IO tool can extract from the HTML tags of the website. The development of review classification combined with commenter profiles is as follows: 1) Tokenization: cutting by gaps, (end of sentence) or using a dictionary; 2) Stop word removal: removing the necessary word such as prepositions, pronouns, conjunctions and nouns; 3) Stemming: finding the original word of the word; 4) Feature extraction extracting: the characteristics of the text, such as bags of word tolerance, text representation with word frequency (Term Frequency: TF) and text representation with inverse frequency (Term Frequency – Inverse Document Frequency: TF-IDF); 5) Feature selection using

information gain or the characteristics of the domain of interest; 6) Classification using the machine learning, grouping, etc; 7) Measurement of research efficiency analyzing by Precision, Recall, F-measure, Accuracy, etc.

3. Research Methodology. As seen in Figure 1, this research has compiled the hotel opinions framework, a form of text (Text) that hotel guests review based on their experience. It consists of various opinions such as positive and negative comments on the feature of interest. These guest reviews can tell how you feel or the attitude of the guests towards the hotel. We collected 12,644 hotel guest reviews from 30 hotels with hotel descriptions from www.tripadvisor.com using the Import.io APIs to collect hotel guest reviews. Normally the hotel descriptions represent the concept or advantage of the hotel from the hotelier. The descriptions are in the text and unstructure format. In this

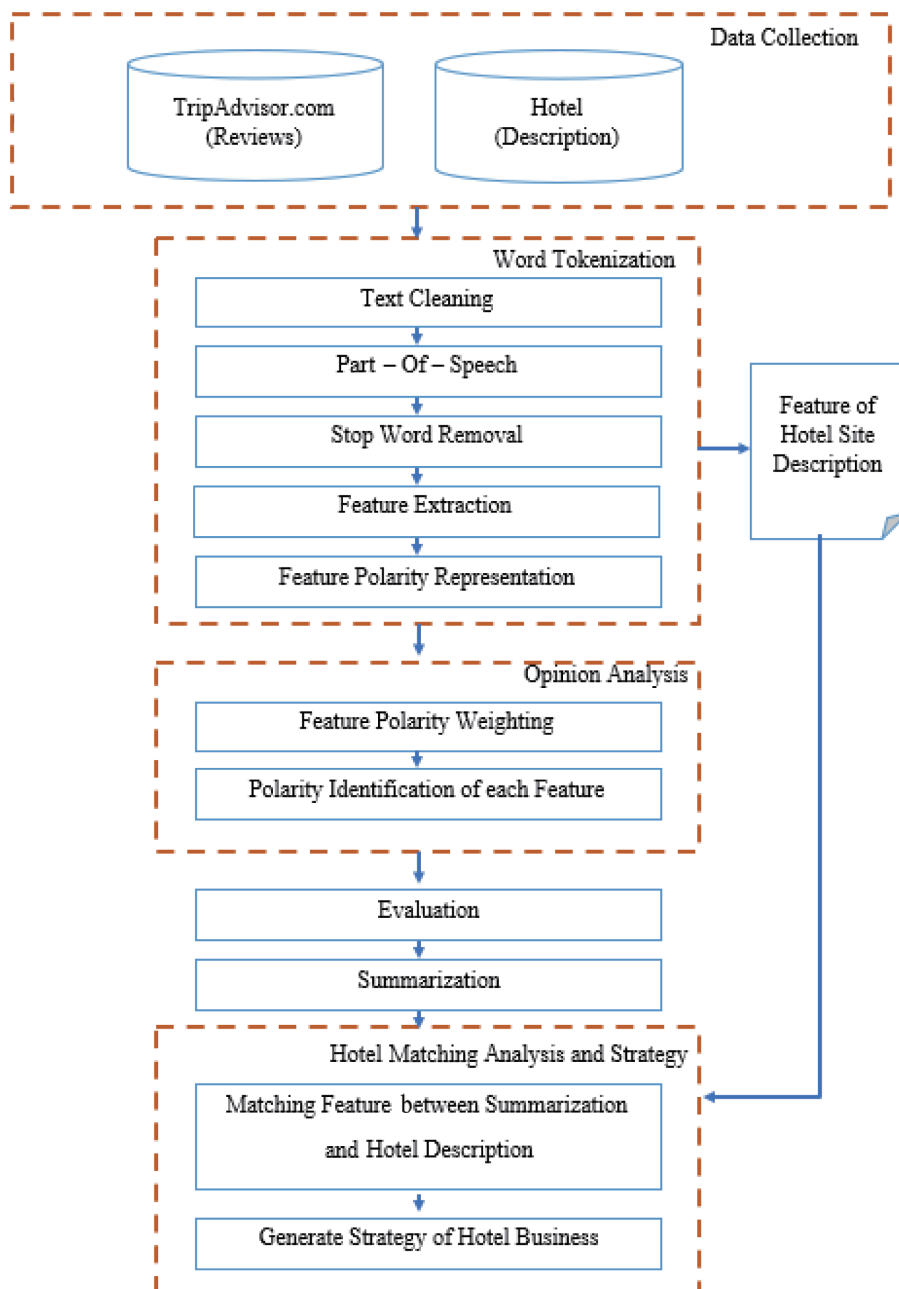


FIGURE 1. The proposed framework of hotel customer opinions for business

research, only the opinions written in English were selected and used in the analysis. After this process, the words have tagged the POS tagging type (Part-of Speech Tagging). After removing the stop word successfully, this research extracts the characteristic words and polarized words to select key traits used to identify traits and polarizations to analyze feelings of interest [15-19]. It consists of location (Location), rooms (Rooms), service (Service), food and beverages (Food) and room prices (Price) by these features, collected from the previous research [15-19]. In addition to these words, there are also synonyms. To find synonyms use the Synonym dictionary. After extracting the attributes from hotel guest comments, this research extracts the term functional opinion polarity and chooses the words that act as verbs, adjectives (ADJ), and adverbs (ADV) because these types of words can express feelings in which words appear in several parts of SentiWordNet depended on the context and function of the word. In this research, the terms in each context are analyzed together. If the word appears more positive, it will conclude that the term is positive, and it is negative. For example, Good has 27 words that appear on SentiWordNet. There are 23 positive, three neutral, and one negative meaning. Therefore, it concluded that Good is a word that expresses positive feelings. After the data has passed the word segmentation process, the word will analyze and summarize the opinions of each feature as positive or negative. The hotel descriptions, in text, are fed to this process as same as guest review. By examining the number of expressive context words, it appeared in front of each feature word; it can be written in Figure 2.

```

For  $i$  in List Feature : List Feature  $\leftarrow$  [Location, Service, Food, Room, Price]
If found word in feature  $i$  :
    PolarityPositive  $\leftarrow \sum_{i=1}^N$  (PolarityPositive);
    PolarityNegative  $\leftarrow \sum_{i=1}^N$  (PolarityNegative);
    IF(PolarityPositive > PolarityNegative);{
        Feature $_i$   $\leftarrow$  Positive;
    } ELSE IF (PolarityPositive = PolarityNegative){
        IF (BackWord of Polarity = Polarity) {
            Feature $_i$   $\leftarrow$  Positive ;
        } ELSE {
            Feature $_i$   $\leftarrow$  Negative;
        }
    } ELSE {
        Feature $_i$   $\leftarrow$  Negative;
    }
}

```

FIGURE 2. The polarity rule assigning

After completing the opinion analysis, the research compares the results between the experiment and the experts to measure hotel guest reviews analysis effectiveness. This research measures the efficiency of opinion analysis by considering the Precision, the Recall, and the overall efficiency (F-measure), as shown in Equations (1)-(3).

$$\text{Precision} = \frac{n_a}{N_p} \quad (1)$$

$$\text{Recall} = \frac{n_a}{N} \quad (2)$$

where n_a is a number of corrected prediction, N_p is a total number of prediction and N is the total number of reviews.

$$\text{F-measure} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (3)$$

In step hotel matching analysis and strategy, to summarize the results of the opinions hotel guest reviews analysis and hotel description, this research prepared a graph showing the correct polarity of the correctly predicted five features of interest. In the next step, the main contribution of the proposed method compares the hotel guest reviews with the hotel description. Is it consistent with the hotel’s description in 5 features: Location, Service, Food, Rooms, and Price? And how many agreeable? If the comments are in the same direction as the hotel description, then the hotel’s description corresponds to the reviews of hotel guests, as shown in Figure 3.

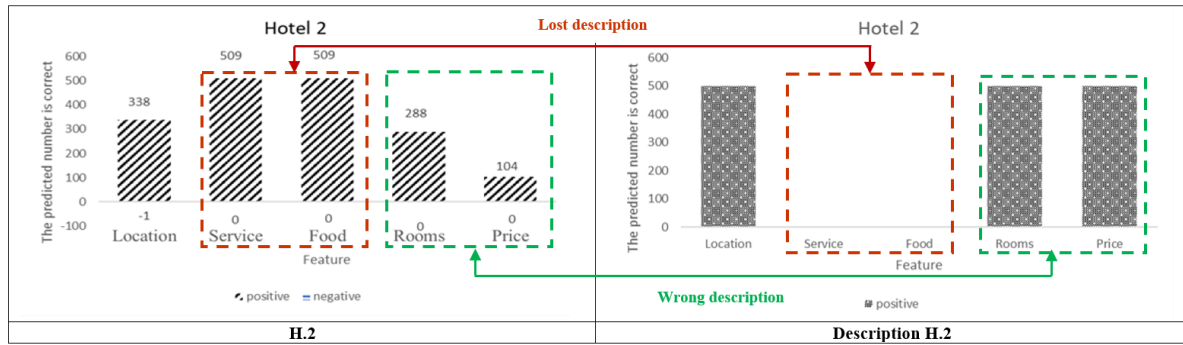


FIGURE 3. The comparison of guest reviews with the 2nd hotel description

As seen in Figure 3, on the left, guests have positive opinions on Service, Food, Location, Rooms and Price, respectively, and on the right, the hotel has written a description of its strengths, which has been extracted into 3 areas: Location, Rooms and Price. In the description of the hotel’s strengths, it does not mention Service, Food, which are both aspects of the hotel, and the guests have the most positive reviews.

4. Evaluation Results.

4.1. Data collection. In this research, the opinions of hotels located in Bangkok, Thailand, were collected from 30 hotels with quality levels ranging from 2 to 5 stars. The collected data were aggregated to specify by five experts to determine each comment labeled to positive or negative interesting features. The experts considered feature words together with synonyms for five feature words: Location, Service, Food, Rooms, and Price. In the polarity of each feature, experts look at the definitions of positive and negative sentiments. After the five experts had already identified the answer, this research selected only the opinions which more than three experts labeled the same opinion. After collecting data successfully and selecting the comments written in English, the word’s function is labeled using the Penn Treebank Tagset [20] and deleted the stop word using a dictionary-based technique, and searches for all the words with synonyms for feature words using a dictionary of synonyms that the researcher has compiled from other studies [15-19]. If a word is synonymous with a feature word, the word will replace with the given feature word. This simplifies the grouping of feature words and can be analyzed more efficiently. For example, the word “City” will replace with “Location” feature.

After substituting synonyms for all feature words, subsequent processes substitute the polarity words. This research uses SentiWordNet word corpus [21] by selecting only the prefixed terms JJ (Adjective, comparative), JJR (Adjective, superlative), JJS (Adjective, base form), VB (Verb, base form), VBD (Verb, past tense), VBG (Verb, gerund), VBN (Verb, past), VBP (Verb, non-3d), VBZ (Verb 3d), RB (Adverb), RBR (Adverb, comparative) and RBS (Adverb, superlative), and analyzes the value of polarity words by considering the score appearing on SentiWordNet. When the sum of sentiment on each word is calculated and compared to identify the polarity, if there were more positive

words than negative words, the words were positive substituting with a plus sign, and else negative represented with minus sign.

4.2. Results. Since there are many polarity words in each comment sentence, it consists of positive and negative words. Therefore, in some comments, a polarity word appears after the feature word. These polarities are summed by the value. For example, the feature word “Rooms” has a preceding score number of 9. The positive words are represented by a + in the front of 10 of the feature word “Rooms”, and the negative words are represented by a – in the preceding one, then sum the expression. After that, the feature polarity volume is the maximum comparison between the before and after volume of the feature. There is a method of selection if any polarity word appears in front of any feature word [18]. That polarity word is considered to belong to that feature word. After analyzing the guest opinion terminals towards each feature, the results of the experiments were compared with the expert’s results in order of data. For efficiency measures in this research Accuracy, Recall, and F-measure measurement methods were used. The results obtained from the experiment are shown in Table 1.

TABLE 1. Performance measurement (Precision, Recall, and F-measure)

Hotel name	Total number of comments	The correct predicted number	Total predicted number	Total number of results answer	Precision	Recall	F-measure
H.1	694	1,613	2,248	2,186	0.69	0.82	0.75
H.2	682	1,749	2,001	2,136	0.85	0.81	0.83
H.3	700	788	1,442	2,006	0.82	0.81	0.81
H.4	707	1,477	1,613	1,847	0.74	0.79	0.76
H.5	294	524	686	821	0.76	0.63	0.69
H.6	397	470	663	795	0.70	0.59	0.64
H.7	373	397	600	687	0.64	0.56	0.6
H.8	393	792	1,200	1,094	0.66	0.72	0.69
H.9	303	713	1,023	927	0.69	0.76	0.72
H.10	365	597	862	920	0.69	0.64	0.66
H.11	386	522	1,030	753	0.50	0.69	0.58
H.12	376	583	906	944	0.64	0.61	0.62
H.13	399	593	1,006	1,015	0.58	0.58	0.58
H.14	395	533	861	1,031	0.62	0.51	0.56
H.15	285	347	581	730	0.59	0.47	0.52
H.16	285	348	581	675	0.59	0.51	0.55
H.17	399	535	775	820	0.69	0.65	0.67
H.18	413	479	932	1,318	0.51	0.36	0.42
H.19	400	575	732	930	0.78	0.61	0.68
H.20	405	415	1,046	843	0.39	0.49	0.43
H.21	376	808	1,230	1,559	0.65	0.51	0.57
H.22	420	350	680	1,000	0.51	0.35	0.42
H.23	400	317	1,003	686	0.31	0.46	0.37
H.24	403	329	1,009	700	0.32	0.47	0.38
H.25	398	222	876	736	0.25	0.30	0.27
H.26	389	419	1,054	798	0.39	0.52	0.45
H.27	384	376	1,035	778	0.36	0.48	0.41
H.28	383	451	907	820	0.49	0.55	0.52
H.29	400	328	1,104	780	0.29	0.42	0.34
H.30	396	417	965	848	0.43	0.49	0.46
Average	420	602.23	1,021.7	1039.43	0.57	0.57	0.56

From Table 1, the efficacy results of the analysis of hotel guest reviews with the highest values were H.2. H.2 is the 2nd hotel from 682 comments, 1,749 correct predictions, 2,001 total predictions, 2,136 total number of results answer, 85% accuracy, efficiency by maximum 83%, the highest values 82% of Recall were H.1, and on the other hand, the efficacy results of the analysis of hotel guest reviews with the lowest values were H.25. Details of H.2 and H.25 are shown in Figure 4.

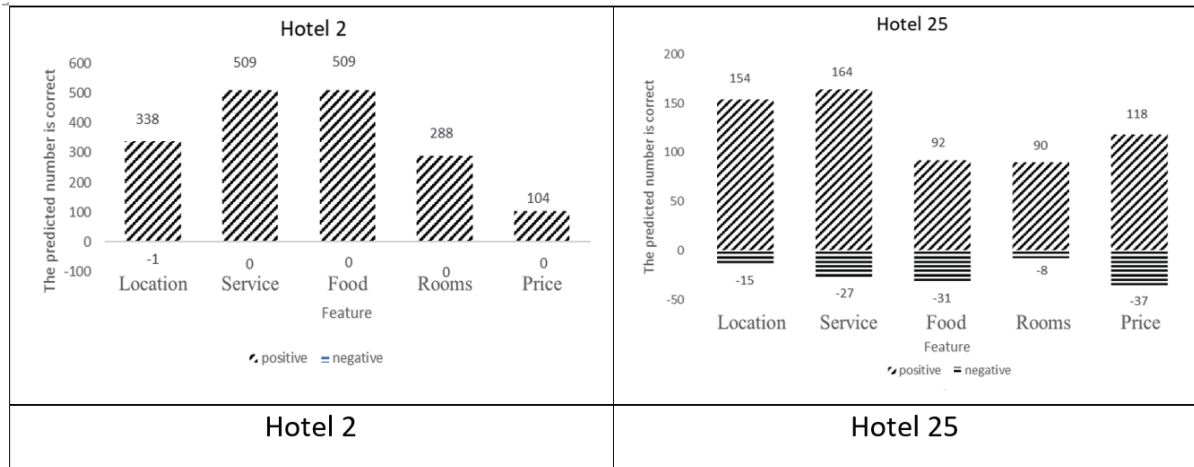


FIGURE 4. The feature of hotel 2 and hotel 25

As seen in Figure 4, the results (Reviews of Location, Service, Food, Rooms and Price) show that H.2 has more positive reviews than H.25. As a result, the F-measure of H.2 (0.83) is higher value than H.25 (0.27).

4.3. Comparison of hotel guest reviews with hotel descriptions. This research compared the opinions between the hotel reviews and the hotel description, and the example results will show in Figures 5 and 6.

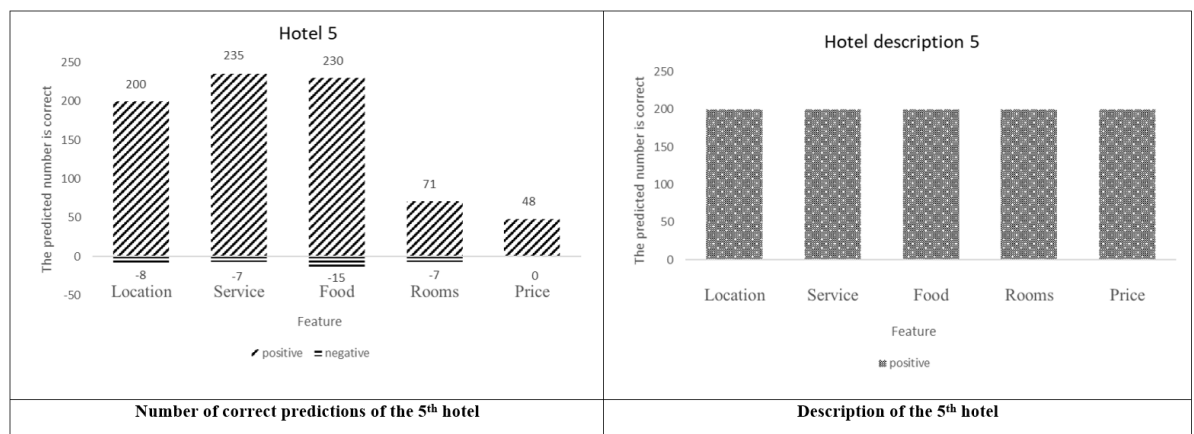


FIGURE 5. A comparison of the guest reviews and the hotel description of the 5th hotel

As seen in Figure 5, guests' reviews were very positive for Service, Food, and Location, respectively. The hotel has written a description of the strengths of all five features. Describe the results of the comparison between the correct predictions and the hotel description as follows: On the left, the majority of guests leave a large number of positive reviews of the three features: Service, Food, Location, and fewer of the positive reviews of the Rooms and Price features. Therefore, the business owner should improve or create a new strategy to improve the Rooms and Price features. Additionally, all four features,



FIGURE 6. A comparison of the guest reviews with the 14th hotel description

Service, Food, Location, and Rooms, had negative reviews. Business owners should improve or be interested in these four service features to meet guests' needs and build loyalty to the hotel.

As seen in Figure 6, the results can be explained as follows from the picture on the left. It was found that the guests gave positive feedback on the features of Service, Food, and Location, respectively. The hotel wrote a description of one aspect of the hotel's strengths, "Location"; thus, the hotel description should cover all features from the positive guest reviews. This is because some guests may wish to receive other hotel services not mentioned by this hotel. Emphasis is placed on Service and Food characteristics to guests looking for features that meet their stay decisions.

In our research, it is different from another works, for example, [22] proposed a systematic approach that first constructs classifiers to identify helpful reviews and then classifies the sentences in the helpful reviews into 6 hotel features. [7] proposed the analysis and prediction rating from customer reviews who commented as open opinion using probability's classifier model, and [8] proposed a method for categorizing the opinions of hotel users. with text mining methods and Latent Dirichlet Allocation (LDA) methods. In our work, the results are completely compared with the description of the hotel.

5. Conclusion. This paper proposed the comparison between the correct prediction results and the hotel description, it was found that business owners should advertise through a hotel description, particular a distinctive feature or a feature with a high number of positive guest reviews. Moreover, it probably improves the service in each aspect of the hotel to meet the guests' needs. Looking for the real reasons why guests leave negative reviews to improve the service makes the guest impression and loyalty to the hotel, and then advertise. The hotel management team can use the research findings as a guideline for the hotel's marketing strategy development. In our research, it is different from another works such as [7,8,22]. In our work, the results are completely compared with the description of the hotel that is useful for hotel ecosystem such as customer and hotelier.

In the future, classification will be enhanced by adding classes to distinguish distinct differences such as strong positive, ordinary positive, ordinary negative and strong negative. In addition, this research uses features from other research studies which may not be able to solve problems in the hotel business comprehensively. Therefore, if we listen to the sound of the customer experience in the hotel by considering the word frequency, key features, hotel management will use this research for findings as a guideline for developing hotel marketing strategies more effectively.

REFERENCES

- [1] S. F. King and T. F. Burgess, Understanding success and failure in customer relationship management, *Industrial Marketing Management*, vol.37, no.4, pp.421-431, 2008.
- [2] B. Milović, Social media and eCRM as a prerequisite for hotel success, *Management Information System*, vol.7, no.3, pp.26-31, 2012.
- [3] A. Nair, R. Khasgiwala and S. Mishra, Improving customer relationship management using data mining, *Int. J. Sci. Eng. Res.*, vol.7, no.2, pp.1161-1163, 2016.
- [4] M. Sarkar, S. Roy, A. Das and A. Nath, Recognition of handwritten characters using binarization, image segmentation and correlation matching technique, *International Journal of Advance Research in Computer Science and Management Studies*, vol.5, no.6, 2017.
- [5] M. R. Yaakub, Y. Li and Y. Feng, Integration of opinion into customer analysis model, *2011 IEEE 8th International Conference on e-Business Engineering*, pp.90-95, 2011.
- [6] A. Kongthong, N. Angkawattanawit, C. Sangkeettrakarn, P. Palingoon and C. Haruechaiyasak, Using an opinion mining approach to exploit web content in order to improve customer relationship management, *Proc. of PICMET'10: Technology Management for Global Economic Growth*, pp.1-6, 2010.
- [7] W. Songpan, The analysis and prediction of customer review rating using opinion mining, *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, pp.71-77, 2017.
- [8] A. C. Calheiros, S. Moro and P. Rita, Sentiment classification of consumer-generated online reviews using topic modeling, *Journal of Hospitality Marketing & Management*, vol.26, no.7, pp.675-693, 2017.
- [9] F. Z. Ennaji, A. El Fazziki, M. Sadgal and D. Benslimane, Social intelligence framework: Extracting and analyzing opinions for social CRM, *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, pp.1-7, 2015.
- [10] M. Afzaal and M. Usman, A novel framework for aspect-based opinion classification for tourist places, *2015 10th International Conference on Digital Information Management (ICDIM)*, pp.1-9, 2015.
- [11] A. A. Mohammed and B. Rashid, Customer Relationship Management (CRM) in hotel industry: A framework proposal on the relationship among CRM dimensions, marketing capabilities, and hotel performance, *International Review of Management and Marketing*, vol.2, no.4, pp.220-230, 2012.
- [12] C. Bucur, Using opinion mining techniques in tourism, *Procedia Economics and Finance*, vol.23, pp.1666-1673, 2015.
- [13] B. Agarwal, N. Mittal, P. Bansal and S. Garg, Sentiment analysis using common-sense and context information, *Computational Intelligence and Neuroscience*, vol.25, 2015.
- [14] A. Iriani, Hendry, D. H. F. Manongga and R.-C. Chen, Mining public opinion on radicalism in social media via sentiment analysis, *International Journal of Innovative Computing, Information and Control*, vol.16, no.5, pp.1787-1800, 2020.
- [15] M. Eirinaki, S. Pisal and J. Singh, Feature-based opinion mining and ranking, *Journal of Computer and System Sciences*, vol.78, no.4, pp.1175-1184, 2012.
- [16] P. P. Tribhuvan, S. Bhirud and A. P. Tribhuvan, A peer review of feature based opinion mining and summarization, *International Journal of Computer Science and Information Technologies (IJCSIT)*, vol.5, no.1, pp.247-250, 2014.
- [17] A. Levi, O. Mokryn, C. Diot and N. Taft, Finding a needle in a haystack of reviews: Cold start context-based hotel recommender system, *Proc. of the 6th ACM Conference on Recommender Systems*, pp.115-122, 2012.
- [18] A. Kongthong, C. Haruechaiyasak, C. Sangkeettrakarn, P. Palingoon and W. Wunnasri, HotelOpinion: An opinion mining system on hotel reviews in Thailand, *Proc. of PICMET'11: Technology Management in the Energy Smart World*, pp.1-6, 2011.
- [19] B. Ray, A. Garain and R. Sarkar, An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews, *Applied Soft Computing*, DOI: 10.1016/j.asoc.2020.106935, 2021.
- [20] M. P. Marcus, B. Santorini and M. A. Marcinkiewicz, Building a large annotated corpus of English: The Penn Treebank, *Computational Linguistics*, vol.19, no.2, pp.313-330, 1993.
- [21] B. Ohana and B. Tierney, Sentiment classification of reviews using SentiWordNet, *Proc. of the 9th IT&T*, 2009.
- [22] C.-F. Tsai, K. Chen, Y.-H. Hu and W.-K. Chen, Improving text summarization of online hotel reviews with review helpfulness and sentiment, *Tourism Management*, vol.80, 104122, DOI: 10.1016/j.tourman.2020.104122, 2020.