African Journal of Hospitality, Tourism and Leisure, Volume 9 (2) - (2020) ISSN: 2223-814X

Tourist sentiment analysis on TripAdvisor using text mining: A case study using hotels in Ubud, Bali

Nasa Zata Dina*
Faculty of Vocational Studies
Universitas Airlangga
Surabaya, Indonesia
ORCID ID: 0000-0002-8560-8626
E-mail: nasazatadina@vokasi.unair.ac.id

Corresponding author*

Abstract

This study analyzes and extracts the TripAdvisor customers' experiences from the user-generated data as the topic of interest from online review customers in TripAdvisor application. The aims of this study were to illustrate on how customers' reviews on social media are able to be used as an evaluation and visual modeling tool. The online review customers are processed using sentiment analysis and text mining. Findings show that customers often review their experience where they have stayed based on their last stay period. The Decision Tree Algorithm is better to classify the sentiment analysis result than the Naïve Bayes Algorithm in the field of accuracy. However, in the field of precision and recall, Naïve Bayes Algorithm is often better than the Decision Tree Algorithm. The text mining results reveal that TripAdvisor customers tend to use words such as "night", "pool", and "time" in negative sentiments expressed after or during a stay.

Keywords: Hospitality, sentiment analysis, social media analytics, TripAdvisor, visual analytics

Introduction

Nowadays, information diffusion has been obtained from the internet. People often express their opinion towards services and products via well-known social media such as Twitter, Facebook, Amazon, and TripAdvisor. This opinion becomes an important source of information. If this resource is well managed, it will be beneficial for the field of business and marketing efforts (Zhang, Li & Chen, 2012). The information is likely to have an impact on the number of room reservations at a hotel (Zhao, Ye & Zhu, 2016). On the other hand, the information can be used by the management of a hotel to escalate inter alia service quality, and staff service provision which are critical considerations. (Rhee, Yang & Kim, 2016; Nicolaides, 2008, 2012).

The latest studies about the analysis of hotel reviews are quite diverse based on social media which has been used as the resource. Xiang et al (2015), obtained a large number of consumer reviews from Expedia.com (Xiang, Schwartz, Gerdes & Uysal, 2015). They constructed a model from hotel guest experience and related it to the satisfaction rating. Meanwhile, Jurek et al and also He et al (2015) used twitter posts to examine the sentiment analysis on each post (Jurek, Mulvenna & Bi, 2015; He, Wu, Yan, Akula & Shen, 2015; Alita, Priyanta & Rokhman, 2019). In 2017, Xiang et al., did a comparative analysis from major social media platforms in the fields of hospitality and tourism. The major social media platforms that have been discussed are TripAdvisor, Expedia, and Yelp (Xiang, Du, Ma & Fan, 2017). Also, some researchers focused on a part of a region. For example, He et al in 2017 only focused on online reviews from hotel customers in China (He, Tian, Tao, Zhang, Yan & Akula, 2017) and Chang et al in 2017 used online reviews from a hotel in the United States of America (Chang, Ku & Chen, 2017).

This research involved multidisciplinary area such as tourism and hospitality, computer science and information systems. The attributes which affects TripAdvisor users' experiences were investigated by using users' online reviews of those who were staying in Ubud, Bali. Text mining and sentiment analysis were applied in this study to help analyzing the online review. Sentiment analysis was used to divide between positive and negative reviews from the guests at each hotel through their opinions. While text mining is used to identify the key attributes, the key then indicates the keyword which often appears based on the number of occurrences from both negative and positive reviews.

Methodology

The objectives were to analyze the sentiment analysis of public review toward their experiment staying in the hotel and also to mine the frequent words which appear on the public reviews. In this study, the researcher applied a case study in Ubud, Bali, and used TripAdvisor as the data source. The system architecture is shown in Figure 1. It is divided into five stages: (1) Data extraction; (2) Data pre-processing; (3) Sentiment classification; (4) Sentiment analysis result; and (5) Performance evaluation and comparison.

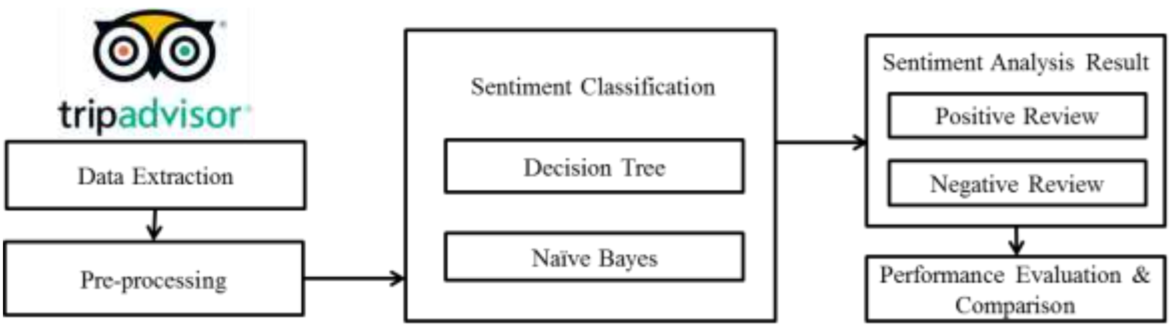


Figure 1. System architecture

Figure 2. Sample positive review from TripAdvisor

Figure 3. Sample negative review from TripAdvisor

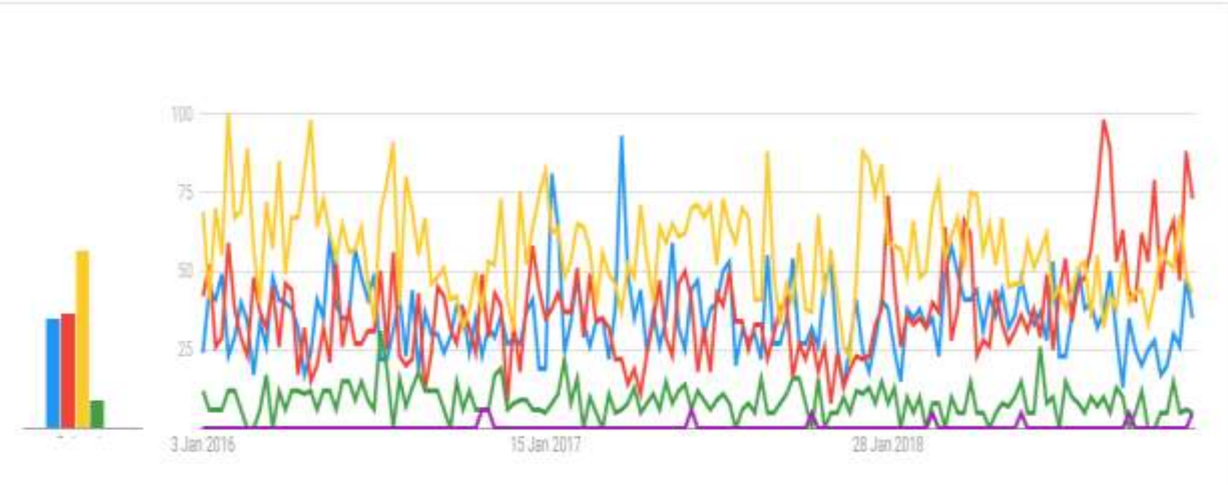


Figure 4. Google trends of 5 selected hotel

Data extraction

The data used for this study were extracted from the TripAdvisor website. The data taken consisted of a hotel's name, ratings, title and review. Figure 2 and Figure 3 are the sample reviews from the TripAdvisor website. The data extraction is specifically for hotels in Ubud, Bali. There are five hotels selected in this study. There are two reasons for selecting those five hotels: (1) the hotels are the most well-known hotel in Ubud so that the hotels' name are the most often searched for. Figure 4 shows the search result on Google trends from the five hotels; (2) the popular hotels have more users' reviews than the unpopular one (Valdivia, Hrabova, Chaturvedi, Luzon, Troiano, Cambria & Herrera, 2019). The data review from each hotel was limited to a period from January 2016 until July 2019.

Data preprocessing

Preprocessing task was applied to the online review's data: (1) define the sentiment of the review based on the rating starts where 1-3 rating stars for negative sentiment and 4-5 rating stars for positive sentiment; (2) convert all the capital letters into lower letters; (3) erase punctuation, number, and stop words; (4) use Porter Stemmer (Porter, 1980) to stem each word. After doing preprocessing, there were only 972 reviews left.

Sentiment classification algorithms

The Decision Tree algorithm is included in supervised learning algorithms. The main aim in using a Decision Tree algorithm is to generate a training model so that it will be used to predict which class the testing data will belong to. The Decision Tree has its root to start predicting the class label and then it jumps to the next branch of the root based on which value the data has followed (Leonard, 2017). The second supervised learning algorithm is Naïve Bayes. The Naive Bayes algorithm predicts the model based on past experience, so it is known as the Bayes Theorem. The Naïve Bayes has the characteristic such as it has a very strong assumption of independence from each condition. One of the advantages of using Naïve Bayes is that it only requires a small amount of training data to determine the estimation of the parameters that are needed while in the classification process (Davies, 2018). After data pre-processing stage, the data were classified into two categories which were positive and negative. From a total of 972 reviews, the data were split using three ratios; they were 50:50, 66:34, and 80:20 as summarized in Table 1.

Table 1. Training and testing data

Algorithm	Ratio		
	50:50	66:34	80:20
Training data	486	642	778
Testing data	486	330	194
Total	972	972	972

The experiment was conducted using the Decision Tree and Naïve Bayes classification algorithms. Both algorithms are supervised learning (Shobha, Rangaswamy, 2018). Supervised learning provides a predicted model; the input and the output have been known. The prediction is generated from the model, so that the classification model has been learned. An algorithm from supervised learning will train the model to obtain the prediction from new data. The new data is often called as testing data. The examples of supervised learning are linear regression, logistic regression, and neural networks, apart from the Decision Tree, Support Vector Machine (SVM), Random Forest, Naïve Bayes, and K-Nearest Neighbor (Shobha, Rangaswamy, 2018; Talabis, McPherson, Miyamoto, Martin & Kaye, 2015). Besides supervised learning, there is also an unsupervised learning. It is the opposite of supervised learning, there are not labels on the data because the training data does not yet exist so that the data has not been classified yet. The algorithms on the unsupervised learning are usually more complex than supervised learning's algorithms because the algorithms are more complex so the processes are more time-consuming. The examples of well-known unsupervised learning are Clustering and Principal Components Analysis (Talabis, McPherson, Miyamoto, Martin & Kaye, 2015).

African Journal of Hospitality, Tourism and Leisure, Volume 9 (2) - (2020) ISSN: 2223-814X

In this study, the supervised learning approach was used. It was because there was a dataset of public reviews from TripAdvisor. The dataset was classified based on the label. The labels were negative or positive. The labeling process of the review dataset had been done in the data pre-processing. The analytic tool used in this study is RapidMiner. RapidMiner is an application for designing the process of data analysis without any programming language required (Ristoski, Bizer & Paulheim, 2015). From this result, there would be a suggestion between two algorithms and which algorithm works best for data of hotel reviews.

Sentiment analysis result

The identification process of sentiment is associated with the data text (Antoniou, Dimitriou & Pereira, 2019). In this study, the result of the sentiment analysis describes how accurate the sentiment analysis predicted by the algorithms is. Before that, the sentiment analysis of each text in TripAdvisor was determined by annotators. The negative and positive analyses have been used as a sentiment analysis.

Performance evaluation and comparison

The measurement for evaluation on classification algorithm are precision, accuracy and recall. The three measurements were calculated based on confusion matrix. The confusion matrix is drawn in Table 2. The formula of the accuracy, precision and recall are written in formula (1)-(3). Accuracy was obtained by dividing the correctly classified data by the total classified data. While precision is the division between correctly classified data from a class and predicted data. Last, recall was the division of correctly predicted data of a class and all the actual data from its class (Padmaja, Sameen & Fatima, 2013).

$$\text{Accuracy} = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} * 100\% \tag{1}$$

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive} * 100\% \tag{2}$$

$$\text{Accuracy} = \frac{True\ Positive}{True\ Positive + False\ Negative} * 100\% \tag{3}$$

Table 2. Confusion Matrix

Class	True Negative	True Positive
Pred. Negative	True Negative	False Positive

Text Mining was used in this study to identify the frequent words appearing in the review so that the topic which was commonly spoken of in a negative and positive review were identified. The summary of the number of words appearance was the result of text mining. The large number of textual data was extracted to obtain “high quality” information, because it served as a big help in providing more insights. Text clustering, data extraction, sentiment analysis are the result of text mining (Talabis, McPherson, Miyamoto, Martin & Kaye, 2015). The term occurrences were used as vector creation so that the number of the term could be calculated.

Results and discussion

Sentiment analysis classification results: Performance evaluation and comparison

The performance of sentiment analysis classification results is listed in Table 3. It shows that the highest accuracy was obtained by the Decision Tree algorithm (91.77%). The ratio used was 50:50 which means that the number of training data and testing data were on the same ratio while the algorithm was tested. The highest percentage for precision was obtained by the Naïve Bayes algorithm (99.45%) on 66:34 ratio. For the recall aspect, the highest recall was obtained by Naïve Bayes by 98.28% on 66:34 ratio.

Table 3. Accuracy, Precision, Recall

Algorithm	Accuracy (%)			Precision (%)			Recall (%)		
	50:50	66:34	80:20	50:50	66:34	80:20	50:50	66:34	80:20
Decision Tree	**91.77**	90.61	90.72	93.90	92.28	94.38	96.25	96.69	94.38
Naive Bayes	71.40	72.42	75.26	99.25	**99.45**	99.12	97.67	**98.28**	97.06

Text mining result

The majority of sentiment analysis results in this study are positive. The guests showed a lot of positive aspects of their experiences, and the text mining result is shown in Table 4. For example, the number of occurrences of “stay” in positive sentiment is 452. It means that the word “stay” appears in a positive sentiment in as many as 452 words. Otherwise, the number of occurrences of the word “hotel” in negative sentiment is 365. It means that there are 365 words of “hotel” with negative sentiment connotations.

Table 4. Text mining results

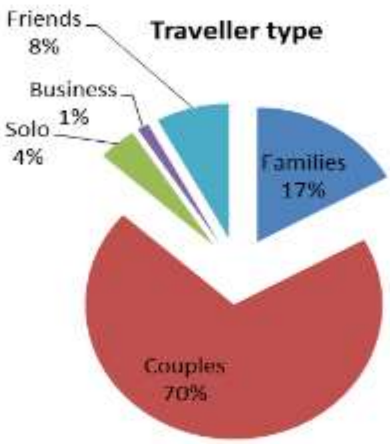
Positive		Negative	
Word	Count	Word	Count
stay	452	hotel	365
hotel	379	room	304
staff	311	stay	300
villa	292	villa	224
service	224	service	198
resort	207	staff	183
room	181	night	156
amazing	179	pool	152
place	161	time	149
night	155	resort	146
time	141	food	129
beautiful	137	good	121
food	133	breakfast	113
experiment	126	nice	113
pool	126	restaurant	106
great	125	great	104
view	124	experience	102
arrive	119	place	97
love	94	check	96
restaurant	94	want	96
welcome	89	expect	95
perfect	85	manager	92
good	81	area	87
honeymoon	79	book	87
location	79	price	87
friendly	73	arrival	84
make	73	Bali	83
wonderful	73	guest	80
breakfast	71	star	78
visit	66	thing	75
help	64	view	68
river	64	beautiful	67
book	63	look	66
moment	63	review	65

African Journal of Hospitality, Tourism and Leisure, Volume 9 (2) - (2020) ISSN: 2223-814X

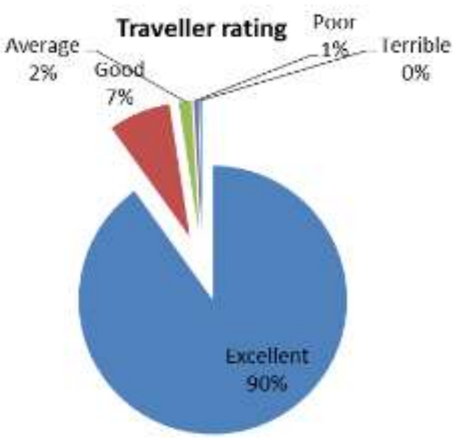
excellence	62	offer	64
expect	61	disappoint	61
rice	61	told	60
nice	60	love	59
trip	60	make	58
check	59	call	57
manager	56	know	57
recommend	56	hour	56
butler	55	feel	55
made	55	come	53
incredible	54	made	52
luxury	54	property	51
private	54	take	51
property	54	people	50
enjoy	53	walk	50
star	53	river	48
want	52	season	48
come	51	treatment	47
world	51	bathroom	46
attention	50	morn	46
review	50	clean	45

Visual analytics using Tableau and Word cloud

The analysis was made based on the timeline when the guests stay. There were two categories processed and this can be seen on the pie chart in Figure 5, the pie chart is made using Tableau software. There are traveler type and traveler ratings. Traveler type includes couples, families, friends, business, and solo. Figure 5 (a) indicates that the majority travellers in Ubud, Bali are Couples and Families. Both Couples and Families often give reviews as opposed to other types of travelers. The couple's traveler type has the biggest portion of the pie table - 70% and it was followed by the family's traveler type at 17%. The main reason is because Ubud has a lot of Bali's arts scene, museum and galleries which is a popular destination for couple and family trips. Figure 5. (b) shows that the traveler rating for five hotels, for the most part, is an "excellent" rating. The number of reviews that have an excellent rating is 90%. By performing a similar analysis, the management of each hotel could identify travellers' type and the information of the hotel's rating so that the services could be enhanced. In addition, the marketing strategies could be adjusted. In 2015, Torres et al., linked the number of reviews based on the popularity and its rating (Antoniou, Dimitriou & Pereira, 2019). Thus, that is the reason why this study only included the five most popular hotels in Ubud, Bali. The reviews are easy to obtain and further analysis are easy to conduct.



(a)



(b)

African Journal of Hospitality, Tourism and Leisure, Volume 9 (2) - (2020) ISSN: 2223-814X

Figure 5. (a) Traveller type; (b) Traveller rating

To present the result of the text mining, word clouds is used. It is shown in Figure 6 to visualize the aspect-sentiment keywords. The word cloud was made based on Table 4. The word itself is the result of text mining. These words were divided into two categories: positive and negative. More detailed information about how the words are extracted is already given in text mining. Figure 6 (b) shows the word cloud for a positive review. This word cloud's result could be used as support to the hotel management as decision-makers for their marketing strategies. In contrast, Figure 6 (a) indicates the word cloud for negative review while the size of the word indicates the number of their appearance. Therefore, the bigger word, the most frequent the word appears in the review.

The most frequent keywords in both negative and positive reviews are "stay", "service", "room". This indicates that the traveler observes more to the service and the amenities of the room. Moreover, the traveler often give a high rating on "staff", "villa", "hotel", "resort", "food". In contrast, lower rating on "night", "pool", "time" as seen in Figure 6.(a). The example of a suggestion for the improvement of the facilities is that a hotel can cooperate with a favorite traditional restaurant in Bali to avoid food complaints and also hold interesting events at night.

a).Word cloud negative review

b). Word cloud positive review

Figure 6. Word cloud from the review

Conclusion

This paper discusses a more detailed analysis and visualization of user reviews from TripAdvisor. This study used ranking data, sentiments, and types of tourists in different hotels but in the same area. This shows that hotel review ratings are not the only source of sufficient data to find the original value of hotel review ratings. However, the study that has been done is largely limited to sentence level analysis. The main contribution of this study is how the integrated work process for collecting and processing data, extracting and classifying information is then supplemented with visualization of TripAdvisor data. Visual analytics is presented using word clouds.

References

Al-Amrani, Y., Lazaar, M. & Elkadiri, K. E. (2017). Sentiment Analysis using supervised classification algorithms', *In Proceedings of the 2nd international Conference on Big Data, Cloud and Applications,* ACM, 61.

Alita, D., Priyanta, S. & Rokhman, N. (2019). Analysis of Emoticon and Sarcasm Effect on Sentiment Analysis of Indonesia Language on Twitter, *Journal of Information Systems Engineering and Business Intelligence*, 5(2), 100-109.

Antoniou, C., Dimitriou, L. & Pereira, F. (2019). *Mobility Patterns, Big Data and Transport Analytics*. Amsterdam: Elsevier.

Chang, Y. C., Ku, C. H. & Chen, C. H. (2019). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor, *International Journal of Information Management*, 48, 263-279.

Davies, E.R. (2018). *Computer Vision.* Fifth Edition. Amsterdam: Elsevier.

He, W., Tian, X., Tao, R., Zhang, W., Yan, G. & Akula, V. (2017). Application of social media analytics: a case of analyzing online hotel reviews, *Online Information Review*, 41(7), 921-935.

He, W., Wu, H., Yan, G., Akula, V. & Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks, *Information & Management*, 52(7), 801-812.

Jurek, A., Mulvenna, M. D. & Bi, Y. (2015). Improved lexicon-based sentiment analysis for social media analytics, *Security Informatics*, 4(1), 1-13.

Leonard, L.C. (2017). *Advances in Computers*. Volume 105. Cambridge: Academic Press.

Nicolaides, A. (2012). Service quality provision in upmarket restaurants: a survey of diners in three restaurants in a Gauteng casino complex. *African Journal of Hospitality, Tourism and Leisure*, 2(2).

Nicolaides, A. (2008). Service quality, empowerment and ethics in the South African Hospitality and Tourism industry and the road ahead using ISO 9000/1, PhD thesis, University of Zululand: South Africa.

Nisbet, R., Elder, J. & Miner, G. (2009). *Handbook of Statistical Analysis and Data Mining Applications*. Burlington: Elsevier.

Padmaja, S., Sameen S. & Fatima. (2013). ‘Opinion Mining and Sentiment Analysis–An Assessment of Peoples’ Belief: A Survey’, *International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC),* 4(1), 21-33.

Porter, M. (1980). An algorithm for suffix stripping, *Program*, 14(3), 130-137.

Rhee, H. T., Yang, S.-B. & Kim, K. (2016). Exploring the comparative salience of restaurant attributes: A conjoint analysis approach, *International Journal of Information Management*, 36, 1360-1370.

Ristoski, P., Bizer, C. & Paulheim, H. (2015). Mining the Web of Linked Data with RapidMiner, *Web Semantics: Science, Services and Agents on the World Wide Web*, 35, 142-151.

Shobha, G. & Rangaswamy, S. (2018). *Handbook of statistics: computational analysis and understanding of natural languages: principles, methods and applications*. Amsterdam: Elsevier.

Talabis, M. R. M., McPherson, R., Miyamoto, I., Martin, J. L. & Kaye, D. (2015). *Information security analytics: analytics defined.* Amsterdam: Elsevier.

Valdivia, A., Hrabova, E., Chaturvedi, I., Luzon, M.V., Troiano, L., Cambria, E. & Herrera, F. (2019). Inconsistencies on TripAdvisor reviews: A unified index between users and Sentiment Analysis Methods, *Neurocomputing*, 353, 3-16.

Xiang, Z., Du, Q., Ma, Y. & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism, *Tourism Management*, 58, 51-65.

Xiang, Z., Schwartz, Z., Gerdes Jr, J. H. & Uysal, M. (2015). ‘What can big data and text analytics tell us about hotel guest experience and satisfaction?’, *International Journal of Hospitality Management*, 44, 120-130.

Zhang, Z., Li, X. & Chen, Y. (2012) Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews, *ACM Transactions on Management Information Systems (TMIS)*, 3(1), 5.

Zhao, D., Ye, Q., & Zhu, K. (2016). The influenceof consumers’ information search behavior on purchase decisions for experience goods: Empirical evidence from hotel industry, *WHICEB 2016 proceedings*, p297-302.