Analysis on Customer Satisfaction Dimensions in Peer-to-Peer Accommodation using Latent Dirichlet Allocation: A Case Study of Airbnb

Kevin M. Situmorang., Achmad N. Hidayanto., Alfan F. Wicaksono., Arlisa Yuliawati
Faculty of Computer Science – Universitas Indonesia
kevin.mahendra@ui.ac.id, nizar@cs.ui.ac.id, alfan@cs.ui.ac.id, arlisa@cs.ui.ac.id

***Abstract*— Customer satisfaction becomes a key influencer for people's habits or daily activities. One of the examples is in the decision-making process about whether they will use specific products or services. People often need other's review or rating about what they are going to use or consume. In this research, by using customer's online review that available from Airbnb website, we try to extract what are the most talked factors about peer-to-peer accommodation, and how customer sentiment about them. We use Latent Dirichlet Allocation (LDA) to extract that factors and conduct sentiment analysis by utilizing semantic analyzer from Google Cloud NLP. We analyze which factors that has more effect on customer satisfaction, not only in general but more specific based on customer gender and tourism destination object. The result shows that factors related to social benefit and service quality have impact on customer satisfaction, moreover different customer gender and different tourism object destination bring different sentiment among customer. We also find several factors that can be improved by the owner of the accommodation to improve customer satisfaction toward their services.**

***Keywords*—Customer Satisfaction, Sentiment Analysis, Peer-to-Peer Accommodation, Latent Dirichlet Allocation**

I. INTRODUCTION

SHARING economy is a peer-to-peer activity to get, to give, or to share things or services through online platform [1] [2]. Collaborative consumption which is based on the concept of sharing economy has brought the society into a new level of consumption pattern. One of the instance of this business model is peer-to-peer accommodation which known as the introduction of Airbnb in 2008 [3]. It is said that peer-to-peer accommodation offering another solution for travelers who want unique experience in choosing accommodation. Physically, Airbnb does not own any assets, which are the accommodations (room, house, hotel, etc.). Airbnb becomes an intermediary between the owners of the accommodations and the customers. The accommodations registered in Airbnb website are available for rent by customers and the process basically goes like the offline accommodation reservation.

Customer or user satisfaction is defined as subjective evaluation from customer or user whether specific product or service have satisfied their expectation [4]. This evaluation is important for others who are willing to use specific services or products. For example, in the field of peer-to-peer accommodation, before deciding to stay in a specific place, people need to know how well the service quality of both the accommodation and the staff are. Therefore, good experience will bring positive word-of-mouth that will take effect on the increase of company sales and profits [5]. Since customer satisfaction is beneficial for all stakeholders in this peer-to-peer accommodation case, it is important to identify in what factors customers satisfied with certain products or services.

Previous research found that *service quality*, *staff hospitality*, *reputation of the accommodation*, and *security* become the affecting factors for customer satisfaction, specifically related to peer-to-peer accommodation [4]. Additional factors such as *enjoyment*, *social benefits*, *economic benefits*, *sustainability*, *amenities*, and *location benefits* are also identified to be important [2] [3] [6] [7] [8]. Generally, these previous researches used quantitative, qualitative, or combination of using questionnaire and focus group to gather what are important factors behind customer satisfaction. In this method, to avoid useful information are not captured from the questionnaire, we need to involve expert in the construction of each question [9].

In the era of social media, User Generated Content (UGC) as an expression of products or services are useful for companies to know customer's acceptance of their products [10]. In the case of Airbnb platform, this information can be obtained from customer's online review. Realize that trust factor also affecting customer's decision, Airbnb managed to provide only useful reviews and ratings from previous customers through their policy, i.e. (1) deleting reviews that are not represent to customer's personal experience, (2) reviews that are promised by specific reward, and (3) reviews that are driven by the threat of extortion [11]. Airbnb also describe their extortion policy to prevent the creation of any reviews that violates three points above. By utilizing this relevant review data, our main goal is to automate the process of identifying the affecting factors for customer satisfaction, specifically in the domain of peer-to-peer accommodation.

As mentioned in previous research about customer review data [9], customers may not explicitly mention specific dimensions of customer satisfaction in their review. Instead, they indirectly describe other indicators or attribute represent those dimensions. By employed Latent Dirichlet Allocation (LDA) as topic modeling method, this research

This work was supported by Universitas Indonesia under PITTA Grant 2018, with the contract number 1862/UN2.R3.1/HKP.05.00/2018

was able to extract hidden dimension affecting visitor satisfaction towards online hotel reviews that were not found by using traditional method. We adopt this method to extract hidden topics from Airbnb's customer review data as the affecting factors for customer satisfaction. Generally, our experiment tries to figure out the answer of these questions:

- What are the factors or dimensions that affecting customer satisfaction in peer-to-peer accommodation?
- How better are the extracted factors compared to previously identified (in previous research)?
- What is the heterogeneity of perceptions for different groups of customers and different groups of tourism destination object?
- What are the most affecting factors for customer satisfaction based on regression analysis?
- What is the relationship between online rating provided by Airbnb with the customer satisfaction?

We organize the rest of this paper as follows: Section II explains the literature review, continued by research method in Section III. Section IV describes the result and discussion, then the last section brings the conclusion.

II. Literature Review

A. Extract Hidden Topics through Latent Dirichlet Allocation

In this research, we want to automatically extract that factors by utilizing Latent Dirichlet Allocation (LDA) as one of techniques for identifying hidden topics from unstructured data. This technique is an efficient unsupervised technique since it can handle not only large data but also periodically and sparse data [12]. By the assumption that each document is made up of topics with certain proportion, LDA tries to find the topics stored in a document.

Here is the outline of topic modeling using LDA [9] [12]:

1. A corpus is a collection of *m* documents represented in $C = \{D_1, D_2, D_3, .. D_m\}$. Each document D_d consist of *n* terms represented in $D_d = (w_1, w_2, w_3, .. w_n)$. There are *p* topics, represented in $T = \{T_1, T_2, T_3, ... T_p\}$.
2. LDA step by step:
   a. A document-term matrix is decomposed into two matrices: M_1 which is the document-topic matrix and M_2 which is the topic-term matrix.
   b. Each term then is randomly assigned to a topic. The assignment will be updated based on the multiplication of P_1 and P_2 until it reaches certain convergence limits. P_1 and P_2 respectively are:
      - P_1 or P(topic T | document D) is the proportion of terms in a document assigned to a topic T.
      - P_2 or P(word w | topic T) is the proportion of topic T to all documents D derived from term w.
   c. In LDA, parameter α represents the document-topic density while β parameter represents the topic-term density. The higher value of parameter α indicates that the higher number of topics are construct a document (and vice versa). The higher value of parameter β indicates that a topic is composed of more terms (and vice versa).
3. Other important issue in LDA technique is to determine the number of topic to be extracted [13]. We use harmonic mean method to estimate marginal likelihood of each LDA model, as done by [14].

B. *Sentiment Analysis* using Google NLP API

Sentiment analysis or opinion mining is a process of analyzing opinion, sentiment, evaluation, reaction, and emotion towards specific person, organization, issue, or event [10]. As mentioned in [10], sentiment analysis can be differentiated in several levels, they are:

1. *Document level analysis.* This level of analysis classifies documents into positive, negative, or neutral sentiment based on analysis of the whole content of document that talks about specific entity or product, not about comparison of entities or products.
2. *Sentence level analysis*. In this level, we analyze whether a sentence belongs to positive, negative, neutral, or does not belongs to any sentiment.
3. *Aspect or entity level analysis* (*feature-based opinion mining* [15]). The analysis is conducted towards a specific opinion which is consists of opinion target and sentiment of related opinion. Opinion target can be a specific entity and/or specific aspect being talked in that opinion. For example, the sentence "*The iPhone's call quality is good, but its battery life is short*" has "*call quality*" and "*battery life*" as its opinion targets. And evaluation "*good*" and "*short*" for each mentioned target respectively.

After obtaining the list of hidden topics using LDA from the review document collection as the opinion targets, we try to classify each topic into positive, negative, or neutral sentiment as the evaluation of the targets. We utilize Google Cloud NLP API[1] to get the sentiment score. The API provides several methods for text analysis and annotation, one of them is *Sentiment Analysis* method. This method provides sentiment score of a passage that indicates the passage's overall emotion. The sentiment score is ranged between -1.0 (negative sentiment) to 1.0 (positive sentiment). For example, a sentence "*bathroom outside is incredible!*" scored 0.6 and has positive sentiment by using this Google API's method.

III. Research Method

A. Data Collection and Text Preprocessing

We use Scrapy[2] combined with Airbnb's API to gather customer review data from Airbnb website. Since Airbnb does not provide gender information, we also use Genderize[3] to identify gender of each customers. We need to

[1] https://cloud.google.com/natural-language/docs/basics#interpreting_sentiment_analysis_values

[2] Python based framework for extracting data from website (https://scrapy.org/)

[3] This tool determine gender based on the first name. (https://genderize.io/)

know whether a customer is male or female to analyze customer satisfaction based on gender.

We filter the whole data in several steps as follows:

1. Choose the review data which gender is identified. We obtain 32,072 reviews i.e. 45.73% reviews from male customer and 54.27% reviews from female customer.
2. Choose review in English since about 88.22% reviews are written in English. We obtain 27,711 reviews.
3. Choose review that associated with tourism object destination based on the accommodation distance from tourism object (no more than 30 km). Our final data is 24,911 reviews.

We then do several preprocessing steps to transform our textual data into numerical data by following these steps.

1. Replace numerical text that refers to currency with '*price-indicator*' notation, for example Rp 8000, 150k, $50.
2. Tokenization based on whitespace character to get list of single terms or tokens.
3. Part-of-Speech Tagging to identify the class of each term, e.g. noun, verb, adjective, etc.
4. Obtain base form of each term through lemmatization to reduce the number of unique terms obtained from the whole review data.
5. Named Entity Recognition using NLTK[4] for identifying the kind of entity owned by each term. There are nine provided entities, they are *organization*, *person*, *location*, *date*, *time*, *money*, *percent*, *facility*, *gpe* [16]. We add one more identifier to represent people's name that refers to name of the accommodation's host. The identifier is '*pname-indicator*'
6. Noun extraction by only selecting terms that belongs to noun class since they more representing a specific topic compared to those that belong to other classes.

After these steps, we construct bag-of-words for each document, containing 14,263 terms/words. Each document is represented into a vector $D[(A_1, B_1), \dots (A_m, B_m)]$ where m is the number of terms (14,263), A_i is identity number of terms i, and B_i is frequency of term i appears in document D.

B. Topic Extraction using LDA

We use Gensim[5] to build LDA model and get numbers of topics from Airbnb's given customer review data. One of the parameters needed to build LDA model is the number of topic. We do experiment on 31 number of topics, started from 20 to 50 and choose the best option. For each number of topics, we build its LDA model and evaluate the harmonic mean value of its log likelihood. From this experiment, the highest value is -1,946,092, reached when we use 43 topics.

From LDA model, each of 43 topics is represented by several most contributing words together with their proportion into related topic, not explicitly by the name of topics. For example, Table I shows list of words representing certain topic with proportion of each word. By investigating the logical relationship of all contributing words, we try to find a specific topic name for each topic. We also recheck thoroughly into sample of documents to see the usability or context of the words in documents. For example, topic #18 is named as "*friendly host*". There are eight topics that their name cannot be identified because the contributing words relatively not connected each other. We named these topics under the name *Unidentified #1* to *Unidentified #8.* We also choose six words (as seeds) manually to represent each topic since there are several words that are not related into topic they represent. The bold and italic words in Table I shows the selected seeds for topic "*friendly host*".

Each document is associated to topics by considering the probability of each topic in each document trough the seeds. We choose some topics for each document which probability exceed a specific threshold to avoid not relevant assignment from a document into specific topics. The threshold value is calculated from the total probability value of each topic in specific document divided by the number of known topics associated with specific documents. From this process, the result shows that the topics which names cannot be identified are relatively have small distribution among the whole documents. This indicates that customers relatively talk less about these topics or dimensions. As the result, we ignore these eight dimensions so there are 35 remaining topics to be analyzed.

C. Analysis of Dimensions based on Statistical Analysis

After obtaining 35 topics, we then evaluate how customer sentiment towards these topics. Generally, we differentiate our analysis on following aspects:

1) Analysis of Sentiment for Each Dimensions

We start our analysis on aspect level by using sentiment analyzer tools provided by Google Cloud NLP to identify the sentiment of each sentences. We choose sentences which contains at least one seed word and run the sentiment analyzer to get the sentiment of these sentences. Overall result using sentiment analyzer from Google Cloud NLP shows that from the collection of sentences containing seeds, 4% of them are identified as neutral sentiment, 6% as negative sentiment and 90% as positive sentiment.

The next step is to determine the sentiment of each dimension through their seeds. We divide sentences based on their related dimension and reanalyze the sentiment score obtained in previous step using Google Cloud NLP API.

TABLE I
EXAMPLE OF TOPIC WITH LIST OF CONTRIBUTING WORDS
(TOPIC #18 / FRIENDLY HOST)

Contributing Word	Proportion (%)	Contributing Word	Proportion (%)
pname-indicator	19	***person***	2.3
minute	8.8	holiday	2
family	8.6	Place	1.8
friend	5.3	year	1.7
time	4.8	company	1.7
people	3.9	contact	1.6
thank	3.6	child	1.6
kind	3.2	clean	1.4
nature	3.1	message	1.3
trip	3.1	lot	1.1

[4] Natural Language Toolkit, a python-based platform for NLP task (https://www.nltk.org/)

[5] Python based library for topic modeling (https://radimrehurek.com/gensim/)

2) Analysis of Dimensions based on Specific Groups

We analyze customer sentiment not only generally towards the whole dimensions but try to explore the heterogeneity of dimensions within the different groups of customers (e.g. female vs male) [9] and added more analysis for different groups of tourism object destination (e.g. sea, culture, and city). Through this step, combined with statistical test, we want to know how significance is the different of male and female customers sentiment towards each dimension. We have the similar objective with the analysis of groups of tourism destination objects, we want to measure how significance is the difference of customer sentiment of sea, culture, and city tourism object towards of the 35 dimensions.

3) Analysis on Online Rating Dimensions

Airbnb website provide two kinds of online rating, they are overall customer rating ranged from 1 to 100 and specific rating for six aspects (*cleanliness, arrival, accuracy, location, value,* and *communication*) ranged from 1 to 10. We run stepwise regression analysis to identify the most significant dimensions affecting customer satisfaction.

We normalize the overall customer rating into 0 to 10 and this value is chosen as the dependent variable, while the rating value for another six aspects as the independent variable. We run stepwise regression based on (1) and (2).

$$y_i = \hat{y} + e_i \quad (1)$$
$$\hat{y} = b_1x_{1i} + b_2x_{2i} + b_3x_{3i} + b_4x_{4i} + b_5x_{5i} + b_6x_{6i} \quad (2)$$

The variable y_i represents overall customer satisfaction, while x_{1i} to x_{6i} represents customer rating towards *cleanliness, accuracy, value, communication, arrival,* and *location* respectively, and e_i represents normally distributed residual error value. The result of this analysis is useful to identify which dimensions provided by Airbnb website that have the most effect on the customer satisfaction.

IV. Results And Discussion

A. The Extracted Dimensions and Their Comparison to Previous Research

Based on the result of topic modeling using LDA, we obtained the total of 35 most discussed topics or dimension among customer review. Our result actually is also agreed upon previous research [6] [7] [8] [9] [17]. This is shown by the result of the calculation of Jaccard Coefficient[6] to get the proportion of overlapping topics obtained by LDA and topics obtained by previous research. The total score is 0.68, so there are more than 50% topics that are also identified as affecting factors in customer satisfaction. Table II shows the comparison of the topics obtained by using LDA with previous research ordered by the proportion or distribution of each dimension towards the whole review document. Furthermore, our experiment using LDA can identify more affecting factors for customer satisfaction compared to previous research result.

Additional result based on Table II, we find that dimensions related to host or staff of the accommodation has relatively high proportion among the whole customer reviews. These dimensions include *host service*, *welcoming host*, *staff service*, and *friendly host*. Generally, they can be categorized as *social benefits* and *service quality*. *Social benefits* also bring significant effect on customer satisfaction since customers also enjoy their experience live with locals [8]. Moreover, most of the *service quality* focuses only on services provided by host instead of service quality of Airbnb platform. Specifically, customer concern mostly about the accommodation itself, including the hospitality and service given by accommodation host or staff rather than about the Airbnb platform. Furthermore, we cannot find any topics related to Airbnb platform from the result of topic modeling using LDA.

TABLE II
Topics Obtained by Using LDA and Obtained from Previous Research (Ordered by Topic's Proportion Towards Documents)

Topic/ Dimension's Name	Previous Research	LDA
Host Service	X	V
Welcoming Host	V	V
House	V	V
Staff Service	X	V
Villa	V	V
Experience	V	V
Transportation Rental	X	V
Friendly Host	V	V
Apartment	V	V
Driver	V	V
Price Value	V	V
Bathroom	V	V
Arrival Experience	X	V
Booking Experience	X	V
Amenities	V	V
Garden	X	V
Beach Vibe	X	V
Hospitality	V	V
Closeness to Shops	V	V
City View	X	V
Easy to Find the Location	V	V
Closeness to Restaurant	V	V
Style and Decoration	X	V
Room Maintenance	V	V
Cultural Experience	V	V
Hangout	V	V
Activity	V	V
Natural Beauty	X	V
Community	V	V
Insider Tip	V	V
Communication	V	V
Public Transportation	X	V
Help Locals	V	V
Bedroom	V	V
Dining	V	V

B. General Customer Sentiment and based on Specific Groups Towards Each Dimension

1) Customer Sentiment Towards Each Dimension

We have already got the result of sentiment analysis of all sentences containing the seeds. Based on the seeds, we classify the sentences based on their dimension to know the sentiment towards each dimension. All dimensions are identified to be indicator of positive word of mouth. There are four dimensions, such as *hospitality, communication, cultural experience, and garden* which reach the highest sentiment score, i.e. 0.7. These dimensions close to *social benefits* aspect, which is in line with the result described in previous subsection that customer's social experience defines their satisfaction.

[6] $Jaccard = \frac{|Topics\ obtained\ by\ LDA \cap Topics\ obtained\ by\ previous\ reseacher|}{|Topics\ obtained\ by\ LDA \cup Topics\ obtained\ by\ previous\ reseacher|}$

2) Customer Sentiment based on Customer's Gender

We start with the identification of the proportion of each dimension toward customer's gender before identifying the sentiment. Through *t*-test[7], the result shows that there are 10 topics that significantly different among male and female customers. They are *cultural experience, experience, host service, price value, staff service, welcoming host, apartment, driver, friendly host,* and *transportation rental.* Since mean value of topic distribution towards male customer is higher than female, it indicates that male customers pay more attention to these factors in determining their satisfaction towards the peer-to-peer accommodation.

We divide the documents into those who authored by male customers and those by female customers then analyze the sentiment for each dimension independently based on customer's gender. All topics get positive sentiment but there are some differences between male and female customer. By using *t*-test, the result shows that there are 9 topics that are identified as significantly different between male and female customer. They are *booking experience*, *host service*, *activity*, *amenities*, *bathroom*, *house*, *transportation rental*, *villa*, *welcoming host*. The *t*-test result also shows the mean value of sentiment on each topic among female customers is higher than the mean value of sentiment on each topic from male customers. This result indicates that sentiments given by female customers are more positive compared to those from male customers.

3) Customer Sentiment based on Tourism Destination Object

In this part, we analyze how the distribution of each topic among sea, culture, and city tourism object. By using ANOVA test we obtain 21 topics that has significant difference between the three categories of the tourism object destination. These topics are *amenities, apartment, experience, friendly host, hospitality, host service, house, price value, staff service, transportation rental, welcoming host, room maintenance, arrival experience, booking experience, bathroom, communication, community, cultural experience, driver, garden, style* and *decoration.* This significant difference indicates that in different tourism destination object, customers talk about different things.

We separate the documents into those who talks about sea, city, and culture tourism destination object then analyze the sentiment of each topic towards these categories independently. There are 11 topics that significantly different among those three categories. They are *arrival experience, driver, host service, house, amenities, experience, transportation rental, booking experience, friendly host, beach vibe, staff service.* All topics are identified as positive sentiment scored from 0.3 to 0.7. From this result we can conclude that in different tourism object, customer have different sentiment towards related dimensions.

C. Analysis on the Relationship between Online Rating with The Sentiment of Extracted Dimension

From the six aspects provided by Airbnb website as the dependent variables (*cleanliness, arrival, accuracy, location, value,* and *communication*), the result of regression

TABLE III
STEPWISE REGRESSION MODEL

Model	R^2	*Adjusted* R^2	*F*	*Sig.*
1[a]	0.5178	0.5175	2014	< 0.001
2[b]	0.6279	0.6275	1582	< 0.001
3[c]	0.6809	0.6804	1333	< 0.001
4[d]	0.6981	0.6974	1083	< 0.001
5[e]	0.702	0.7012	885	< 0.001
6[f]	0.7062	0.7053	749.7	< 0.001

Note:
[a] *Predictors: (Constant), cleanliness, accuracy*
[b] *Predictors: (Constant), cleanliness, accuracy*
[c] *Predictors: (Constant), cleanliness, accuracy, value*
[d] *Predictors: (Constant), cleanliness, accuracy, value, communication*
[e] *Predictors: (Constant), cleanliness, accuracy, value, communication, arrival*
[f] *Predictors: (Constant), cleanliness, accuracy, value, communication, arrival, location*

analysis shows that there is a strong relationship between these variables and the overall customer rating as dependent variable by using significance level 0.1%. Table III shows that the highest R^2 value is obtained by the model which includes all of six dependent variables. This indicates that those six aspect or dimension from Airbnb website also affecting customer satisfaction.

By using Durbin-Watson test (the value is 2.01 and p=0.824), it shows that residual error mentioned in Equation (1) is independent and does not have effect on customer satisfaction. Moreover, *cleanliness* is the most affecting dimension on customer satisfaction since it hast the highest beta coefficient as shown in Table IV. The most affecting aspect for customer satisfaction ordered by their beta coefficient are *cleanliness*, *value, accuracy, communication, arrival,* and *location* respectively, and the regression model for customer satisfaction ($\hat{y}$) is shown in (2).

$$\hat{y} = 0.36 + 0.31 * cleanliness + 0.2 * accuracy + 0.24 * value + 0.1 * communication + 0.09 * arrival + 0.08 * location \quad (3)$$

We also try to correlate the sentiment score we have obtained for each extracted dimension from LDA model with the overall customer rating from Airbnb website. We want to determine whether the customer's sentiment described in review data is as good as Airbnb's overall customer rating. We transform the sentiment result of each dimension into a rating value from 1 to 10 by using (4) [16]:

$$Rating = \left(\frac{P}{P+N} \times 9\right) + 1 \quad (4)$$

Variable P represents the number of positive sentiments, while negative sentiment is represented by N. The rating of overall topics reaches above 8 and more than 50% of the topics reach above 9. The average of this customer

TABLE IV
COEFFICIENT OF DEPENDENT VARIABLE FOR REGRESSION MODEL

Variables	Unstandardized		Standardized Beta Coeff.	t	Sig.
	B	Std. Error			
(Constant)	0.36	0.15		2.43	< 0.05
Cleanliness	0.26	0.02	0.31	17.29	< 0.001
Accuracy	0.18	0.02	0.20	10.38	< 0.001
Value	0.22	0.02	0.24	12.15	< 0.001
Communication	0.12	0.02	0.11	5.90	< 0.001
Arrival	0.10	0.02	0.09	4.96	< 0.001
Location	0.07	0.02	0.08	5.19	< 0.001

[7] We use p value 0.05 and confidence interval 95% for running t-test

satisfaction rating is 93.55, and it almost the same compared to customer rating that available at Airbnb website, which is 93.33.

V. Conclusion

Our study tries to automate the process of extracting dimensions of customer satisfaction in peer-to-peer accommodation domain, specifically based on Airbnb case. The result shows that we can automatically find the dimensions through topic modeling using LDA from the collection of customer review data. There are several undiscovered dimensions that affecting customer satisfaction from previous research that use survey at limited sample.

On the analysis of heterogeneity of perception on different groups of customers, especially based on gender, we find that male customers have more attention toward several factors, while female customers tend to give more positive review than male customer in other factors. In the groups of tourism destination objects, we notice that different tourism object, bring different concern among customers about what factors affecting their satisfaction.

Based on the result of stepwise regression analysis, we know that the whole factors provided by Airbnb online rating significantly affecting customer satisfaction, especially *cleanliness* as the most affecting factor. The accommodation's owner should consider this factor to improve their cleanliness of the accommodation, moreover related to *house, villa, apartment, room maintenance, bedroom,* and *bathroom* since based on the reviews, customers also talk about these factors.

In general, by using *User Generated Content* data such as customer review, it is possible to extract the hidden dimension that affecting customer satisfaction. Through the utilization of NLP method such as topic modeling by using LDA, there will be no more limitation on the scale of the data since this method can handle large amount of data.

VI. References

[1] R. Botsman and R. Rogers, What's Mine Is Yours: The Rise of Collaborative Consumption, New York: Harper Collins Publisher, 2010.

[2] J. Hamari, M. Sjöklint and A. Ukkonen, "The Sharing Economy: Why People Participate in Collaborative Consumption," *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY,* pp. 2047-2059, 2016.

[3] I. P. Tussyadiah and J. Pesonen, "Impacts of peer-to-peer accommodation use on travel patterns," *Journal of travel Research,* pp. 1022-1040, 2016.

[4] R. K. S. Chu and T. Choi, "An importance-performance analysis of hotel selection factors in the Hong Kong hotel industry: a comparison of business and leisure travellers," *Tourism Management,* pp. 363-377, 2000.

[5] W. W. Moe and M. Trusov, "The Value of Social Dynamics in Online Product Rating Frums," *Journal of Marketing Research,* pp. 444-456, 2011.

[6] I. P. Tussyadiah, "An exploratory study on drivers and deterrents of collaborative consumption in travel," in *Information and communication technologies in tourism*, 2015.

[7] M. Möhlmann, "Collaborative consumption: determinants of satisfaction and the likelihood of using a sharing economy option again," *Journal of Consumer Behaviour,* pp. 193-207, 2015.

[8] C. P. Lamberton and R. L. Rose, "When is ours better than mine? A framework for understanding and altering participation in commercial sharing systems," *Journal of Marketing,* pp. 109-125, 2012.

[9] Y. Guo, S. J. Barnes and Q. Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation," *Tourism Management,* pp. 467-483, 2017.

[10] B. Liu, Sentiment Analysis and Opinion Mining, Chicago: Morgan & Claypool Publishers, 2012.

[11] Airbnb, "Apa yang dimaksud dengan Kebijakan Konten Airbnb?," Airbnb, [Online]. Available: https://www.airbnb.co.id/help/article/546/what-is-airbnb-s-content-policy.

[12] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research 3,* pp. 993-1022, 2003.

[13] D. M. Blei and J. D. Lafferty, "Text mining: Classification, clustering, and applications. Topic Models.," pp. 71-94.

[14] T. L. Griffiths and M. Steyvers, "Finding Scientific Topics," *PNAS,* vol. 101 (supl 1), pp. 5228-5235, 2004.

[15] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," in *The tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, Seattle WA, 2004.

[16] S. Bird, E. Klein and E. Loper, Natural Language Processing for Python - Analizing Text with Natural Language Toolkit, O'Reilly Media Inc, 2009.

[17] J. Hamari, M. Sjöklint and A. Ukkonen, "The Sharing Economy: Why People Participate in Collaborative Consumption," *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY,* pp. 2047-2059, 2015.