See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/335946220

Understanding Public Opinion towards New Sharing Economy Business Model Using Content Analysis

Conference Paper · August 2019

DOI: 10.1109/ICIMTech.2019.8843779

| CITATIONS 8 | | READS 117 | |
|----------------|--|--------------|--|
| 3 author | s, including: | | |
| | Andry Alamsyah Telkom University 168 PUBLICATIONS 1,058 CITATIONS SEE PROFILE | 0 | Wachda Yuniar Rochmah Telkom University 1 PUBLICATION 8 CITATIONS SEE PROFILE |

Understanding Public Opinion towards New Sharing Economy Business Model Using Content Analysis

Andry Alamsyah School of Economic and Business Telkom University Bandung, Indonesia andrya@telkomuniversity.ac.id Wachda Yuniar Rochmah School of Economic and Business Telkom University Bandung, Indonesia wachdayuniar@gmail.com Ditya Dwi Adhi Nugroho School of Economic and Business Telkom University Bandung, Indonesia dityaadhi@outlook.com

Abstract-Hospitality, as one of the sectors in the tourism industry, continues to show positive trend every year based on its revenues. In order to understand public opinion, the legacy method such as interview, questionnaire, and other timeconsuming method is still widely used in this sector. However, this method is less efficient compared to data analytics methodology using available data from online source, such as social media Therefore, we need to conduct a research to see a better method in order to understand the public opinion in hospitality sector. There are many methodologies to support content analysis based on unstructured data to understand public opinion. In this research, we use content analysis methodology which consists of Sentiment Analysis, Topic Modeling, and Text Network Analysis to process 483413 obtained tweets from January 7th until March 28th, 2017. The twitter data is gained using Python and R language. The whole process of the research consists of data collection, data preprocessing, and interpreting the results. We use Rstudio, Jupyter Notebook and Gephi softwares during the process. As a case study in the hospitality sector, we use Airbnb as one of the sharing economy business models that allows other people to share their space and begins to be widely used by travelers around the world. Our objective is to understand public opinion towards Airbnb as online-based sharing economy business model which becomes a disruptive innovation for tourism industry. As the result, we are able to apply the combination of content analysis methods to give a good understanding towards the public opinion. Text Network Analysis gives us the ability to summarise large-scale conversation in very fast and real-time fashion. It provides the knowledge by associate most frequent words used in social conversations. Sentiment Analysis shows the feelings and emotion from people about certain topics. Topic Modelling is able to capture popular topics, thus gives comprehensive understanding on how people react towards a new business model

Keywords—sharing economy, content analysis, sentiment analysis, topic modeling, text network analysis

I. INTRODUCTION

Sharing economy allows society to arrange their everyday task into new ways of life such as; drive people around, or share extra rooms. Moreover, another type of sharing economy startup also enables its consumer to share their ideas or information, such as Facebook, Twitter, and other social media. [1]. Sharing economy business gives them more power to people want to utilize their assets. Thus, it provides more efficient way than traditional service. In addition, it also affects the stability of established companies due to its low price. However, they have to face the challenge such as the regulation which is different in each country.

As the case study, Airbnb is an online-base sharing economy that provides various services, including rental. It empowers many people around the world because people are able to monetize their properties Travelers can have access to stay in various kind of places, from a villa to a castle. All of them are brought together in their website and app. The company, established in 2008, has conquered the global market in many countries and becomes one of the big rental places. They provide access to more than 5 million places to stay, spread in 191 countries around the world [2]. Following the democratization sharing information related to a product or service in an online social network, customers share their opinions on social media, including Airbnb customers. Thus, in this research we investigate public opinion towards Airbnb as a new sharing economy business model using pattern exploration based data analytics instead of traditional approaches such as a questionnaire or interview which doesn't give the comprehensive view of the public opinion.

We apply several content analysis methodology which consist of Text Network Analysis, Sentiment Analysis, and Topic Modeling to understand comprehensively customer's perception towards Airbnb as a new sharing economy business. Previous researches have applied these methods in order to understand customer's level of satisfaction to the company or competitors [3] and in a large scale data [4]. As the final result of this research, we are able to identify the topics from the positive and negative opinions and see the relation of the words which are being talked by people.

II. LITERATURE REVIEW

A. Sharing economy

The Sharing economy is a new concept to share the human and physical resource, including the goods and services of different people and organizations. The goal is to allow the sharing and collaboration in all social economic aspects of life, such as collaborative consumption, shared ownership, renting, borrowing, lending, etc. [5]. A collaborative consumption has been expected to lessen problems such as poverty, hyper consumption, or even pollution [6]. Alongside with its advantage, trust has become the main reason of people to engage with sharing economy and distinguished itself from existing rental service [7].

B. Content Analysis

Content analysis is a set of methodologies to analyze written, verbal, or visual communication message [8]. It is

being used with the increasing frequency by a growing array of researchers and gains more attention in the research literature [9]. Content analysis is a method to learn or analyze a communication systematically, objective, and quantitative in order to measure the variable. It is a research method to create a replica and valid conclusion from the context of the data, to get knowledge, new insight, data representation, and be able to be the guide to do an action [10] [11]. As the application to unstructured data, it was conducted to understand how nonprofit organization used Twitter [12], examine the posts relevant to JUUL and youth on the social media [13], and even analyze a famous entertainment topic on YouTube [14]. Content analysis of public opinion had also be conducted in a hospitality industry, describing the Australian tourists' attitudes to new technology in a hotel. However, the research was conducted using a survey [15].

C. Sentiment Analysis

Opinion mining is a text analysis process in order to obtain certain information from a sentence containing opinion [16]. Sentiment Analysis is an intellectual process to extract feelings and emotion from people [17]. In the Sentiment Analysis, the dataset is divided into training-data and testingdata. The dataset division is set to be 70% training-data and 30% of testing-data [18].

D. Topic Modeling

Topic Modeling is an approach to discover topics in documents. One of the algorithm is Latent Dirichlet Allocation (LDA), which is a learning method among words, topics, and documents by assuming certain probabilities which can model a document [19]. Latent Dirichlet Allocation, in the lowest level, has some needed parameters. One of them is the number of topics, which is symbolized by k. α and β are the parameters which define the characteristics of the topic and the distribution of words in each document dan topic [20].



Fig. 1. Latent Dirichlet Allocation (LDA) illustration

 TABLE I.
 THE PARAMETERS OF THE LATENT DIRICHLET ALLOCATION (LDA) MODEL.

| Parameters | Description | | | |
|------------------------|---|--|--|--|
| K | The number of topics | | | |
| Ν | The number of words in document | | | |
| М | The number of document which is going | | | |
| | to be analyzed | | | |
| А | The Dirichlet-prior parameter in each | | | |
| | topic of the document | | | |
| В | The same parameter from the | | | |
| | distribution of word in every topic. | | | |
| φ(k) | The distribution of word for topic k | | | |
| θ(i) | The distribution of topic in document i | | | |
| ω(i,j) | The word j in document i | | | |
| z(i,j) | The assigned topic for (i, j) | | | |
| φ and θ | The Dirichlet distribution | | | |

Fig. 1 is a model which illustrates how Latent Dirichlet Allocation (LDA) works. The black circle represents the variable which is being observed. Latent is the white circle, and the square plays various types of variables. The parameters of this model are described in the table 1 [20].

E. Text Network Analysis

Text Network Analysis is a computer solution to extract network of concepts from texts and understand the meaning behind them [21]. Any text can be represented as a network which consists of nodes and edges. As applications of this method, text network analysis was used to visualize readable graph from texts, to identify properties and its metrics, or even to discern the most influential pathways for the production of meaning within the text [22].

III. RESEARCH METHODOLOGY

We design the research workflow shown in figure 2. The first step is the data collection. We collect Twitter data conversations from the periods of January 7th to March 28th, 2017. We obtained 483413 tweets containing the keyword "Airbnb" by the streaming process, which enables us to gain the real-time data provided by Twitter Application Programming Interface (API).

The second step, data preprocessing, is illustrated in figure 3, which consists of tokenization, stopwords, and stemming. After being pre-processed, the data is processed to execute the main step, which consists of Text Network Analysis, Sentiment Analysis, and Topic Modeling. In the last step, we take the conclusion by interpreting the results.

In the pre-processing step, we prepare the data by cleaning and restructuring the data before analyzing it. The steps we undertook in the pre-processing as follows:

• Tokenization means splitting a text into small parts [23]. Here is the example of the tokenization: "@Airbnb I am very disappointed with Airbnb services". The result is illustrated below,

@airbnb I am very disappointed with Airbnb services

• Stopwords mean the words that appear frequently yet having low value or useless [23]. Here is the preprocessing process with stopwords, using the previous example that has been tokenized: "@Airbnb I am very disappointed with Airbnb services".

@airbnb disappointed Airbnb services

• Stemming is a process to normalize the text into standard language [23]. The example of this process for "@Airbnb I am very disappointed with Airbnb services" is shown below,

@airbnb disappoint Airbnb service







Fig. 3. Preprocessing Process

In a Text Network Analysis, we construct the dominant words to form a network, including the weighted edge result [24]. The more occurrences of a relation between two words, the thicker its edge has. We apply the modularity metric of network analysis to measure a tendency of the words to cluster. In Sentiment Analysis, we categorize the training-test into 3 labels: positive sentiment, negative sentiment, and neutral. The computer learns different kinds of words and its sentiment in testing data from the training data. We use The Naive Bayes Classifier to do the Sentiment Analysis of the data. In Topic Modeling, we infer 3 topics per sentiment using Latent Dirichlet Allocation (LDA).

IV. RESULTS AND ANALYSIS

A. Text Network Analysis

The figure 4(a) shows the first largest group. We are able to see dominant words such as airbnb, tax, and bans. We found several topics. First, a topic about Airbnb that was facing law case during crawling period. Its competitors wanted Airbnb and other online-based companies in the industry to close down since those companies were unable to compete in term of price. They made use of the reason that Airbnb does not pay the taxes, means it does not follow the regulation. Thus, these words show some supports that have been given to Airbnb for helping the traveler to save their trip budget. Second topic is shown by other words such as *illegal*, big, industry, and against. These words are related to the previous topic where some hotel companies contend that Airbnb is an illegal company. In figure 4(b) or the second largest group, it consists of some words such as room, first, and *coupon*. The combination of these words is based on the promo given to the first-time user of Airbnb. The figure 4(d) consists of words such as *trump*, tower, *rented*, and *listing*. These words appear according to people's conversations on Twitter about President Trump's property when he hired his property in March 2017 by Airbnb.



Fig. 4. The top 4 text networks: (a) the first largest group, (b) the second largest group, (c) the third largest group, (d) the fourth largest group.

B. Sentiment Analysis

From the *Sentiment Analysis* process, we obtain 117453 tweets as positive sentiment, 48787 tweets as negative sentiment, and 208001 tweets as neutral. In table 2, the accuracy is 77.70%, which shows how accurate the testing of testing data towards the training data. The kappa score is 66.10% or above 60%, then it is categorized as substantial agreement. The precision score is 77.74%, and recall score is 76.80%. The example of the categorization is given in table 3 below.

TABLE II. THE MODEL SCORE OF SENTIMENT

| CLASSIFICATION. | | | | | | | |
|-----------------|----------|--------|-----------|--------|-----------|--|--|
| No. | Accuracy | Kappa | Precision | Recall | F-Measure | | |
| 1 | 77.70% | 66.10% | 77.74% | 76.80% | 77.26% | | |

TABLE III. THE EXAMPLE OF SENTIMENT ANALYSIS RESULT.

| No. | Text | Sentiment |
|-----|--|-----------|
| 1. | SEC Chief @MichaelPiwowar says everyone | Positive |
| | should be able to invest in startups like Uber and | |
| | Airbnb | |
| 2. | Who's looking for a place to stay in Truckee North | Neutral |
| | Lake Tahoe! | |
| 3. | @FinallyEven @Airbnb I hope more people find | Negative |
| | out what a scam Airbnb is. Destroying property | - |
| | seems to be a common thread w many people. | |

C. Topic Modeling

The result shows that there are two different colors. The blue-colored bar is overall word frequency, while the red one is estimated word frequency in a topic.



Fig. 5. The First(a) and second(b) overall topics about Airbnb

Figure 5 shows us the overall topics about Airbnb using unclassified datasets. As we are able to see in figure 5(a), the most dominant topic in social media about Airbnb is the luxury home/apartment. Many customers of Airbnb talk about their experiences of living in the luxury place when they are using the service provided by Airbnb. In the second topic, shown in figure 5(b), we are able to see some words such as *airbnb, new, travel, weaccept, tips, startup*, etc. In this topic, people mostly talk about the controversy about the closure of Airbnb which is judged to be inappropriate because people argue that Airbnb is an alternative that facilitates travelers to find a place to stay.





Fig. 6. The first (a), second (b) and third (c) positive topics about Airbnb.

The first positive topic which is shown in figure 6(a) mainly tells that there are many people who support Airbnb to keep operating in their country. It is proven by the word such as *weaccept*, *acceptance*, and *starts*. We are also able to see some words such as Chinese, *people*, *good*, *room*, which tells us that many people are in love with the place which is rent by Chinese because of its tidiness. Figure 6(b) above illustrates the second topic of positive sentiment, many people like Airbnb because there are many travelers who feel helped by the presence of new business such as Airbnb, which provide rental service with the better experience than in a hotel. It makes them feel staying in their own house, so the holiday vibe gets better. It is represented by words such as *love*, *new*, *travel*, *like*, *home*, *rent*, *stay*, *fun*, and *hosts*.



Fig. 7. The first (a), second (b) and third (c) negative topics about Airbnb

(a)

The first negative topic in figure 7(a) in this Topic Modeling shows some words such as ban, Trump, travel, Miami, companies, host, housing, fight. The appearance of those words is caused by the protest launched by individual or organization because Airbnb is perceived to harm hotel businesses. On the contrary, there are many people in Miami who feel safe about Airbnb because many foreigners stay around their houses. The second negative topic, shown in figure 7(b), tells us that there are many people who complain about the owner of the place or the provider of the rental place because the place they rent is not suitable for use, such as water tap problem. Furthermore, the price of Airbnb in the city of New York is quite expensive with not-so-good conditions. This is illustrated by the words new, york, city, hosts, illegal, time, house, problem, make, rentals. Figure 7(c) or the third negative topic tells us about the support which is given by people to hotel companies which insist on Airbnb. It is described from the words hotel, suing, big, company, apartment, industry.

Looking back to the results in the whole process, we are able to see the public opinion comprehensively. It is different from the previous research conducted using survey which only shows percentage of positive and negative results of the opinions [15]. Here, we can obtain not only the distribution results of positive and negative opinions, but also the frequent words and its tendencies to connect each other along with the topics inferred in the public opinions.

V. CONCLUSION AND RECOMMENDATION

We have successfully applied large-scale content analysis method which consist of Text Network Analysis, Sentiment Analysis, and Topic Modeling, to understand the public opinion comprehensively. The combination of these methods are suitable for handling public opinion in a fast and appropriate way. As case study about Airbnb, we have uncovered hidden topics for each sentiment on social media Twitter. From the results, we discover that Airbnb as a new sharing economy business model gains more positive sentiments than the negative ones. The positive sentiments show the support to Airbnb to keep operating in spite of being accused by hotel business. At the same time, people with negative sentiments mostly tell that Airbnb may be a threat to their safety and comforts. We have verified the results based on its accuracy and precision, which both of them shows high value.

However, this research uses the classification method called Naive Bayes for machine learning, and a static Topic Modeling. We suggest another text classification method such as Support Vector Machine, K-Nearest Neighborhood, etc. On the contrary of the static method we apply for the Text Network Analysis, Sentiment Analysis, and Topic Modeling,

we suggest dynamic ones so the new sharing economy business model will be able to understand the customer in the period of time. We also suggest further research to apply these methods in another larger scale of data

REFERENCES

- Hamari, J., Sjöklint, M. and Ukkonen A, "The Sharing Economy: Why People Paricipate in Collaborative Consumption," Journal of The Association For Information Science And Technology, 2047–2059, 2015.
- [2] Airbnb, "About Us [Online]," Available: https://press.airbnb.com/about-us/ 2018.
- [3] Alamsyah A. and Saviera F., "A Comparison of Indonesia's E-Commerce Sentiment Analysis for Marketing Intelligence Effort (case study of Bukalapak, Tokopedia and Elevenia)," Sustainable Collaboration in Business, Technology, Information and Innovation (SCBTII) 1(1), 2018.
- [4] Alamsyah A., Rizkika W., Nugroho D. D. A., Renaldi F. and Saadah S., "Dynamic Large Scale Data on Twitter using Sentiment Analysis and Topic Modeling (Case Study: Uber)," 6th International Conference on Information and Communication Technology (ICOICTT), 2018.
- [5] Matofska B., "What is the sharing economy?" *The people who share*, 444, 2014.
- [6] Hamari J., Sjöklint M. and Ukkonen A., "The sharing economy: Why people participate in collaborative consumption," *Journal of the* association for information science and technology, 67(9), 2047-2059, 2016.
- [7] Kim J., Yoon Y. and Zo H., "Why People Participate in the Sharing Economy: A Social Exchange Perspective," PACIS, 76, 2015.
- [8] Cole F.L., "Content analysis process and application Clinical Nurse Specialist," 2(1), 53-57, 1988.
- [9] Neuendorf K. A., "The content analysis guidebook," Sage, 2016.
- [10] Kerlinger F., "Foundations of behavioral research 3rd," New York: Holt, Rinehart, and Winston, 1986.
- [11] Krippendorff K., "Content Analysis: An Introduction to its Methodology," Newbury Park: Sage Publications, 1980.
- [12] Waters R. D. & Jamal J. Y., "Tweet, tweet, tweet: A content analysis of nonprofit organizations' Twitter updates," *Public Relations Review*, 37(3), 321-324, 2011.
- [13] Brett E. I., Stevens E. M., Wagener, T. L., Leavens, E. L., Morgan, T. L., Cotton, W. D. and Hébert, E. T., "A content analysis of JUUL discussions on social media: Using Reddit to understand patterns and perceptions of JUUL use. *Drug and alcohol dependence*", 194, 358-362, 2019.
- [14] Ahmad, U., Zahid, A., Shoaib, M. & AlAmri, A., "HarVis: An integrated social media content analysis framework for YouTube platform," *Information Systems*, 69, 25-39, 2017.
- [15] Dalton, G. J., Lockington, D. A., & Baldock, T. E., "A survey of tourist attitudes to renewable energy supply in Australian hotel accommodation. Renewable Energy," 33(10), 2174–2185, 2008.
- [16] Liu, B., "Web data mining: exploring hyperlinks, contents, and usage data," Springer Science & Business Media, 2007.
- [17] Devika D. M. C., Sunitha and Ganesh A., "Sentiment Analysis: A Comparative Study on Different Approaches," Procedia Computer Science, 87, pp 44-49, 2016.
- [18] Ting, L. S., Ip W. H. and Tsang C. H. A., "Is Naïve Bayes a Good Classifier for Document Classification?" *International Journal of Software Engineering and its Applications*, 5, 37-45, 2011.
- [19] Pasquali and Rodrigo A., "Automatic Coherence Evaluation Applied to Topic Models," Universidade Do Porto Faculdade de Ciencias thesis, 2016.
- [20] Feldman R. and Sanger J., "Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data," New York: Cambridge University Press, 2007.
- [21] Hunter S., "A Novel Method of Network Text Analysis," Open Journal of Modern Linguistics, 4, 350-366, 2014.
- [22] Paranyushkin D., "Identifying the pathways for meaning circulation using text network analysis," *Nodus Labs*, 26, 2011.
- [23] Hofmann M., Chisholm A., Areal N., Buckley P., Carter B., Colton B. and Silipo R., "Text Mining and Visualization: Case Studies Using Open-Source Tools," CRC Press, 2016.
- [24] Alamsyah A., Paryasto M., Putra F.J. and Himmawan R., "Network text analysis to summarize online conversations for marketing intelligence efforts in the telecommunication industry," 4th International Conference on Information and Communication Technology (ICoICT), 2016.