

# OPINION MINING OF TRIPADVISOR REVIEW TOWARDS FIVE-STAR HOTELS IN BANDUNG CITY

Novi Amalia Santika Putri<sup>1</sup>, Andry Alamsyah, S.Si, M.Sc<sup>2</sup>

<sup>1,2</sup>Prodi S1 International ICT Business, Faculty of Economics and Business, Telkom University  
<sup>1</sup>noviamalia94@gmail.com, <sup>2</sup>andrya@telkomuniversity.ac.id

## Abstract

New ways how and where customers look for recommendation and reviews have emerged. People nowadays tend to check the review and compare potential hotels before they go to make sure whether the hotel is suitable with their preferences or not. Travellers' reviews, in the form of User Generated Content (UGC), have created more effective assessment system in comparison to the more traditional classifications such as hotel's review in guidebooks. TripAdvisor as a market leader of review platforms represents the largest collection of UGC. It transforms not only how customers and hotel interact, but also the use of a new approach to knowledge as a source for customer insight. However, it is difficult for web users to read and understand contents due to the large number of reviews. The hotel management also finds it difficult to obtain any review to measure their service quality. This research aims to provide a comprehensive view of service quality for five-star hotels in Bandung city based on UGC. The dataset of online review is obtained from TripAdvisor. The author uses opinion mining technique to decompose user reviews into five dimensions to measure hotel service quality. This technique can help the hotel management to turn mountains of text opinion into fresh customer insights. The result of this research aims to provide an overview of five-star hotels in Bandung city what dimension of service quality that require attention to be improved. This overview can reduce the time required in understanding review contents for both the hotel management and web users. The overview indicates that each hotel has different dimensions that need more attention. Crowne Plaza, GH Universal, Hilton, InterContinental, Padma, Sheraton, and Trans have the least positive result for assurance in comparison to other dimensions, while Aryaduta, Novotel, and Papandayan have the least positive result for tangibles. This can be used for each hotel management to get customer insights that will help them deliver greater value for customers.

Keywords: Service Quality, Multi-class Classification, Opinion Mining, Review Hotel

## 1. Introduction

A service firm such as hotel, can differentiate itself by delivering consistently higher quality than its competitors provide. Unlike product marketers that can differentiate themselves in providing value-added services, or simply excellent customer service. And like product marketers, service providers need to identify what target customers expect in regard to service quality.

In today's world of mobile technology, the growth, significance and future potential of social media are considerable. The second generation of Web-based service (Web 2.0) allows online users to form and participate in social communities to create and distribute Web content [1]. New ways how and where consumers look for recommendation and product reviews have emerged. Post-trip evaluations and travellers' reviews, in the form of User Generated Content (UGC) have created more effective assessment system in comparison to the more traditional classifications, such as hotels' and restaurants' star ratings in guidebooks.

With the rise of Web 2.0 and Social Media, the importance of User Generated Content (UGC) has risen dramatically, according to [2]. Regarding the hospitality industry, online reviews and ratings have gained enormous influence on the performance and success of a hotel. A 1% increase in a hotel's online reputation score leads up to a 0.89% increase in a hotel's average daily rate (ADR), as well as an occupancy increase of up to 0.54% and up to a 1.42% increase in revenue per available room [3].

People nowadays tend to check the review and compare potential hotels before they go to make sure whether the hotel is suitable with their preferences and intentions. TripAdvisor as a market leader of review platforms, represents the largest collection of UGC. It attracts a high number of visitors to its website.

It indicates a new way of communicating that transformed not only how customers and hotels interacted, but also the use of a new approach to knowledge as a source for customer insight. Through the use of this platform, communication become global and real-time. It allows sharing knowledge become simplified as free and inexpensive, as well as easy uploading of information become commonly available.

In this research, the reviews of five-star hotels in Bandung city based on TripAdvisor will be analysed by using opinion mining. Opinion mining which is used to known as sentiment analysis is a part of data mining study which aims to analyse, understand, process, and extract textual data in the form of opinion, sentiment, evaluation and judgement towards an entity such as product, service, organization, etc.[4].

The opinion mining of TripAdvisor review can be analysed by using Naïve Bayes Classifier. It is one of algorithms which is well-known and used to mining the opinion and offering ease of use as well as quick processing time. Besides, it is easy to be implemented because of the structure simplicity yet high level effectivity [5].

## 1. Literature Review

### 2.1 Service Quality

The service-quality model the main requirements for delivering high service quality [6]. It identifies five gaps that prevent successful delivery:

1. Gap between consumer expectation and management perception.  
Management does not always correctly perceive what customers want.
2. Gap between management perception and service delivery.  
Management might correctly perceive customers; wants but not see a performance standard.
3. Gap between service-quality specifications and service delivery.  
Employees might be poorly trained or incapable of or unwilling to meet the standard; they may be held to conflicting standards, such as taking time to listen to customers and serving them fast.
4. Gap between service delivery and external communications.  
Consumer expectations are affected by statements made by company representatives and ads.
5. Gap between perceived and expected service.  
The consumer may misperceive the service quality.

There are five determinants of service quality based on service quality model [6], they are as follows in descending order of importance:

1. Reliability. The ability to perform service dependably and accurately.
2. Responsiveness. The willingness to help customers and provide prompt service.
3. Assurance. The knowledge and courtesy of employees and their ability to convey trust and confidence.
4. Empathy. The provision of caring, individualized attention to customers.
5. Tangibles. The appearance of physical, facilities, equipment, staff, and communication materials.

### 1.2 Opinion Mining

In a business, it will not be separated from the public opinion about the company, it can be opinions about the location of the company, employees, waste produced and manufactured products. Opinion mining or also known as sentiment analysis is often used to measure a product based on the review community, whether giving positive reviews and appreciated by the public, or otherwise. Tracking those opinions may be sought through the website, social media, and online news.

Since the beginning of 2000, sentiment analysis has developed into one of the most active research in language processing. It's also a lot of learning in data mining, web mining, and text mining [4]. In fact, when it has spread from computer science to management science and social science because of its importance for business and society as a whole. Generally there are three levels of opinion mining, they are document level, sentence level, and entity and aspect level [4].

### 1.3 UGC (User Generated Content)

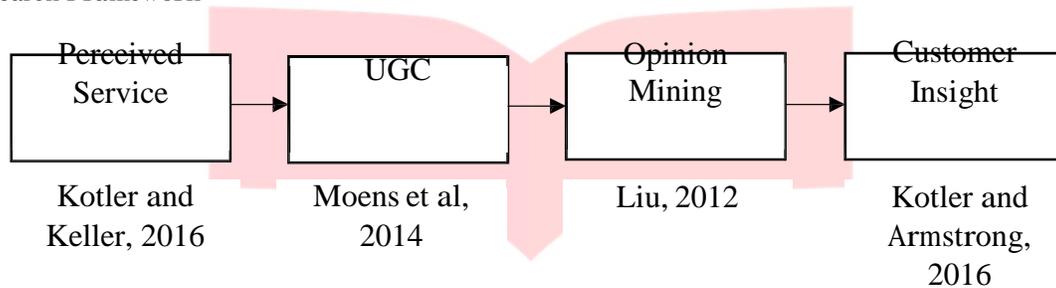
In recent years, the development of the convergence of social networking, mobile computing, and cloud computing is very rapid. This trend has been encouraging users to do most of the social network interaction. Through social networks, users routinely comment on the issue, ask questions, give answers, write about their

views, and make purchases online. Through their gadgets, check-in at your favorite places, and easily share photos and videos, and so forth.

Social media is shared by its users, is associated with metadata, collectively known as UGC (User Generated Content) [7]. UGC is derived from a variety of sources, including social networking sites, such as Facebook and LinkedIn; microblogging sites, such as Twitter; mobile sharing site, such as Path, and Instagram; share information or news sites, such as forums and blogs; image and video sharing sites, such as Flickr and YouTube; and a question and answer sites, such as Wiki-Answers and Yahoo! Answers.

UGC can also be called CGM (Consumer Generated Media), OECD (Organisation for Economic Co-operation and Development) said that UGC as content is published to the site that can be accessed by the public or a social media pages that can be accessed by the user. The sites included in the UGC are blogs, wikis, discussion forums, post, chat, tweet, podcast, pin, digital images, video, audio and various other forms of media created by online users which can be accessed via the web or social media [7].

#### 1.4 Research Framework



#### 2. Data Collection and Method

In this study the source of the data used is based on the concept of UGC (User Generated Content), where the data taken or derived from media websites containing content that is created by the user, both online or offline media. In this case, the media are used as a source of data that social media TripAdvisor. While the content will be taken as a source of data is the comments of the user TripAdvisor in TripAdvisor forum. With the UGC, researchers greatly benefited, because researchers get information and knowledge through collaborative content and contains the opinions [7].

In this study, the data was taken as a commentary on the review in TripAdvisor. Data were taken through the process of web crawling or scrapping using Import.io applications. Before further processed using Rapid Miner applications, the data must first be cleared through pre-processing. Attributes that are not required to be disposed of, until only the remaining data in accordance with the needs of research. With so data presented will be easier to analyse.

Pre-processing need to be done to make high-quality data before step in classification process. High-quality data are the precondition for analysing and using big data and for guaranteeing the value of the data. The data quality standard is composed of five dimensions of data quality, the dimensions are as follows: availability, usability, reliability, relevance, and presentation quality.

Below are the steps being taken in the pre-processing review:

- a) Tokenization is the process of chopping text into small elements and meaningful words.
- b) Stopwords is a process of eliminating frequent words in text documents which are useless.
- c) Stemming is text normalisation process into standard language of text and grammatically correct.

#### 3. Data Analysis and Method

This section explains the data obtained in general to understand research results. The data itself is generated by reviewers in TripAdvisor website in the form of reviews. The review covers five-star hotels in Bandung which are Padma Hotel Bandung, The Trans Luxury Hotel Bandung, Sheraton Bandung Hotel & Towers, Hilton Bandung, The Papandayan, Hotel Aryaduta Bandung, GH Universal Hotel, Crowne Plaza Bandung, Intercontinental Bandung Dago Pakar, and Novotel Bandung. Researcher collected the review on December 13<sup>th</sup> 2016 for each of the hotels.

Classification has been done by using open source RapidMiner 7.3, and the author uses Naïve Bayes as classifier to classify the dataset. From the evaluation process, the author gets 92% precision, 89.67% recall, 90% accuracy, and 0.874 kappa. Just as the category classification, the author also uses open source RapidMiner 7.3, and Naïve Bayes as classifier to classify the dataset. From the evaluation process, the author gets 95% precision, 93% recall, 92.78% accuracy, and 0.857 kappa.

#### 4. Conclusion and Suggestion

Based on analysis and examination which is done by the author in previous chapter, then the conclusions are as follows:

- a) From the precision, recall, accuracy, and kappa we can conclude that Naïve Bayes Classifier can classify both category and sentiment well.
- b) The review of customers can benefit both business and potential guest by collecting information in it.
- c) Each of hotels have their own positive and negative side, from this research they can understand which one can be kept and which one needs to improve.

Based on the testing result and discussion which is done in this research, the author wishes for future researcher to add more data for the training data, so that the result of classification can be more accurate. When the training data has good quality, the result must be better as well.

#### 5. References

- [1] Yoo, K.H., Sigala, M., and Gretzel, U. (2016). Exploring TripAdvisor. Open Tourism: Open Innovation, Crowdsourcing and Co-Creation Challenging the Tourism Industry, 1(2), 239-255. Retrieved from Springer-Verlag Berlin Heidelberg.
- [2] Fritsch, A., and Sigmund, H. (2016). Review Platforms in Hospitality. Open Tourism: Open Innovation, Crowdsourcing and Co-Creation Challenging the Tourism Industry, 1(2), 229-238. Retrieved from Springer-Verlag Berlin Heidelberg.
- [3] Anderson, C. (2012). The Impact of Social Media on Lodging Performance. [online]. Retrieved from <http://scholarship.sha.cornell.edu/chrpubs/5/>
- [4] Liu, Bing. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publisher.
- [5] Hadna, N. M. Shiddieqy., Santosa, P.I., dan Winarno, W. Wahyu. (2016). Studi Literatur Tentang Perbandingan Metode untuk Proses Analisis Sentimen di Twitter. Retrieved from Seminar Nasional Teknologi Informasi dan Komunikasi 2016.
- [6] Kotler, P., and Keller, K.L. (2016). Marketing Management (15<sup>th</sup> edition). Essex, England: Pearson Education Limited.
- [7] Moens, M. F., Li, J., dan Chua, T.-S. (2014). Mining User Generated Content. Boca Raton: CRC Press.