

CHAPTER I INTRODUCTION

1.1 Research Object Overview

1.1.1 TripAdvisor

TripAdvisor was founded February 2000 with its headquarters located in Needham, Massachusetts, United States. It is known as the largest travel site in the world based on comScore Media Metrix 2017. With over 570 million reviews and opinions and 455 million monthly average unique visitors for over 7.3 million accommodations (airlines, attractions, and restaurants) (TripAdvisor LLC, 2017), enables travelers to read and understand the previous traveler's experiences then planning their trips for where to their things.

TripAdvisor is now worth for 6.2 billion dollars. It is employed around 1,400 people with 1.4 billion revenues on 2015. On 2017, TripAdvisor awarded as the America's best midsize companies from Forbes (Forbes Media LLC, 2017).

With the hundred million of reviews that TripAdvisor has, it becomes a favorite place to research. A research from **comScore** shows that the company influences an incredible 40% to 50% of all hotel reservations globally. It shows a significant increase for hotel room sales (The Motley Fool LLC, 2018).



Figure 1. 1 TripAdvisor Logo

Source: TripAdvisor.com

TripAdvisor business model relies on the aggregation of user-generated content (UGC). It attracts visitors to the websites and the revenue streams from advertising and listing fees for other business.

1.2 Research Background

Growth in the number of foreign tourists coming to Indonesia became an activity that benefits Indonesia in various sectors. It gives the advantage in various sectors which helps businesses in tourism industry and also surely increase the country's revenue.

The growth of tourist can be seen from data released by the ministry of tourism which shows an increase every year.



Figure 1. 2 Number of foreign tourist to Indonesia

Source: Author

The graph above is taken from BPS (Badan Pusat Statistik) that shows the number of foreign tourists who came to Indonesia from January 2010 to December 2017. There is a significant increase in the period 2016-2017 that is equal to 21.88% with the number 14,039,799 people.

The data above shows the number of tourists entering Indonesia through all entrances in Indonesia which means all international seaports and international airports. If further examined by looking at the most number per entrance then the largest number is from the I Gusti Ngurah Rai airport located in Bali.

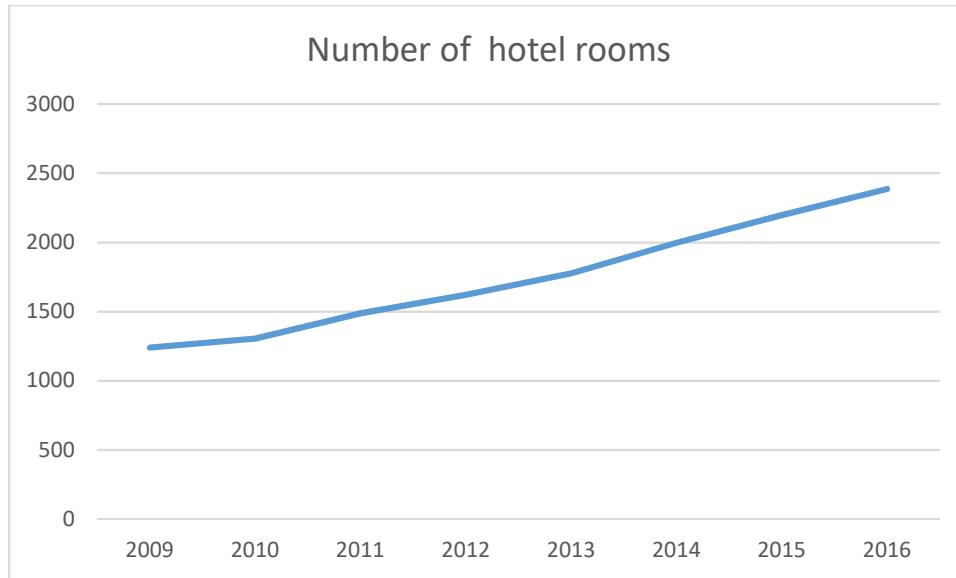


Figure 1. 3 Number of hotel rooms in Indonesia

Source: Author

Due to the large number of foreign tourists who came to Indonesia, this resulted in increasing the number of rooms of hotels and inns in Indonesia that shown in graph above that taken from BPS (Badan Pusat Statistik). The number of starred hotels operating in Indonesia until 2016 is 2,387 hotels with 233,007 rooms.

With the high number of hotel room supplies provided from various hotels then the competition in the hospitality industry will be increased in which hoteliers improved the quality of service and facilities of their hotel. It will be more difficult for customers to make choices. Therefore customers usually tend to check the review on the internet and compare potential hotels before they go to make sure whether the hotel is suitable with their preferences and intentions.

In Marketing 4.0 world, the customer path has already changed. It is used to known as 4A's, now it is changing to 5A's which are advocate is in the last step. In this situation customers become more active in giving their recommendation/advocation of a product or a company whether it is from online media or offline media. Therefore utilization of customers' review become more important for company customer relationship management to know about the complain of their customers, knowing about their brand image/product quality in the perspective of their customers and winning the advocation of their customers in the competition of the industry.

In today's digital world, the activity of checking, reviewing, and comparing of one hotel to another is usually done on the internet. There are many travel forums across the internet that do

the checking, reviewing and comparing the hotels. One of them is TripAdvisor, known as the largest travel site in the world among the traveler (TripAdvisor LLC, 2017). With over millions of reviews of each hotel and a minimum of hundreds per hotel, TripAdvisor is considered more trusted travel site among travelers around the world as it represents to be the largest collection of UGC. In the rise of Web 2.0 and Social Media, the importance of User Generated Content (UGC) has risen dramatically, according to Kaplan & Hainlen (2010) in Fritsch (2016).

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 (Anderson, 2012). Reviews generated by internet users contained in a travel site are regarded as a more credible thing when a prospective hotel customer conducts a hotel survey for himself.

Thus, from the phenomena described above, the hotel management needs insight into what their customers are talking about on the internet to keep relationship with their customers. "*CRM is crucial and has become a niche for firm performance*" (Mohammed & Rashid, 2012). Customer relationship management helps hotels to build relation with customer that in turn creates loyalty, and customer retention. "*Since customer loyalty and revenue are both qualities that affect a company's revenue, CRM is a management strategy that results in increased profits for a business*" (Kulpa, 2017).

In this research, the reviews will be collected from TripAdvisor and will be analyzed using text mining. Text mining which is used to known as sentiment analysis is a part of data mining study which aims to analyze, understand, process, and extract textual data in the form of opinion, sentiment, evaluation and judgement towards an entity such as product, service, organization, etc.

The text mining of TripAdvisor review is using Naïve Bayes Classifier (NBC) algorithm and Latent Dirichlet Allocation visualization because considering several classifiers that we have known, the study by (Ting, Tsang, & Ip, 2011) state that Naïve Bayes has been widely used because of its simplicity in both training and classifying stage. Naïve Bayes method allow each attribute towards the final decision equally and independently from the other attributes. Hence, it is more efficient and more accurate compared to other classifiers (Xhemali, Hinde, & Stone, 2009).

In accordance with the explanation above, the author is interested to do research entitled **“CLASSIFICATION OF HOTEL CUSTOMERS’ REVIEW USING TEXT ANALYSIS FOR CUSTOMER RELATIONSHIP MANAGEMENT”**.

1.3 Problem Statement

In the growth of foreign tourist who came to Indonesia, number of accommodations increases as well. It makes competition among hoteliers high and also for the customers will find it difficult to choose based on their preferences. Therefore, customers tend to compare the options based on reviews online because reviews are considered to be more natural and trusted. Besides the reputation of hotels located in the travel website, influences the average daily rate of hotels (ADR) as well as the hotels revenue (Anderson, 2012).

With large number of reviews, it would be hard for hoteliers to obtain any insights to measure their service quality in a short time regarding with the customer insight. The importance of getting fresh customer insight in a short time for hoteliers will drive to a better output particularly the service and gained more revenue. Such customer insights come from good marketing information (Kotler & Armstrong, Principles of Marketing (16th Edition), 2016). Besides time is a valuable asset in Marketing 4.0 (Kartajaya & Team, 2018). The dynamic change of what customers want is also a challenge. Therefore, bases and customers data movement need to be updated in real-time (Kartajaya & Team, 2018). Such data can be transformed into useful knowledge about them (Kotler, 2002; Lin and Su, 2003; Mguyen et al., 2007; Nasution and Mavondo, 2008; Dev and Olsen, 2000) in (Mohammed & Rashid, 2012).

Gaining deep insights in a short time is a must to know what customer wants and to adapt based on the dynamic changes. This research is expected to assist hoteliers to turn tons on review in TripAdvisor website into customer insights then becomes an insight for service quality in a short time.

1.4 Research Question

1. What is the result of sentiment analysis based on naïve-Bayes classifier?
2. What dimension of service quality need more attention from hoteliers?
3. What is the topic mainly discussed based on the TripAdvisor review?
4. How is the service quality of five-star hotels in Bali based on TripAdvisor review?

1.5 Objectives

1. To know the result of sentiment analysis based on naïve Bayes classifier.
2. To know which dimension of service quality that need more attention from hoteliers.
3. To know what the topic that are mainly discussed based on the TripAdvisor review.
4. To gain insights based on service quality in Bali from TripAdvisor review.

1.6 Significance of Study

1.6.1 Theoretical Aspect

The output of this research is expected to enrich knowledge in the field of content management, specifically in social computing and marketing management especially in customer relationship management which is rapidly evolved along with the development of technology. Some findings hopefully can be reference for further research as well as feedback for academia especially students in terms of gaining knowledge about opinion mining.

1.6.2 Practical Aspect

This research hopefully useful for information in hospitality industry especially hotel business in Indonesia to use TripAdvisor as feedback and evaluation for their business to deliver more value and gained more revenue. As well as being additional references for opinion mining within their business evaluation from the reviewer's point of view.

1.7 Scope of Research

This research is about the tendency of hotel customer behavior to write review about they travel experiences online and the reviews online/reputation online could influences the hotels revenue. This research is devoted to investigate the service quality dimension level of starred hotels in Indonesia by classifying the reviews into the dimension of service quality (servqual). The dimension of service quality (servqual) is based on the research of (Parasuraman, Zeithaml, & Berry, SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality, 1988) "*SERVQUAL is a concise multiple-item scale with good reliability and validity that retailers can use to better understand the service expectations and perceptions of consumers and, as a result, improve service. The instrument has been designed to be applicable across abroad of spectrum of services.*" The dimension of the service quality are Assurance, Empathy, Reliability, Responsiveness and Tangibles but in this research only used 4 of the dimensions (considering that reliability dimensions have no related reviews).

The classifier used is Naïve-Bayes classifier. Considering several classifiers that we have known, the study by (Ting, Tsang, & Ip, 2011) state that Naïve Bayes has been widely used because of its simplicity in both training and classifying stage. Naïve Bayes method allow each attribute towards the final decision equally and independently from the other attributes. Hence, it is more efficient and more accurate compared to other classifiers (Xhemali, Hinde, & Stone, 2009).

For topic modeler Latent Dirichlet Allocation visualization (LDAvis) algorithm is used because based on (Sievert & Shirley, 2014) *”It enables deep inspection of topic-term relationships in an LDA model, while simultaneously providing a global view of the topics, via their prevalence and similarities to each other, in a compact space.”* LDA model itself is *“an example of topic model and is used to classify text in a document to a particular topic. It builds a topic per document model and words per topic model, modeled as Dirichlet distributions”* (Li, 2018).

1.7.1 Location and Research Object

The online platform used in this research is TripAdvisor travel website. It is chosen because considering that it is the online travel site based on a research from **comScore Media Metrix for TripAdvisor site worldwide November 2017** and with approximately 661 million reviews and opinions covering the world's largest selection of travel listings worldwide – covering approximately 7.7 million accommodations, airlines, experiences, and restaurants (TripAdvisor LLC, 2017). The hotel review used in this this research is the review of top-5 starred hotel in Indonesia based on TripAdvisor itself.

The top-5 starred hotel in Indonesia:

- a) Mandapa, A Ritz Carlton Reserve with 492 reviews
- b) Komaneka at Tanggayuda with 431 reviews
- c) Viceroy with 426 reviews
- d) Katamama with 371 reviews
- e) Jamahal, A Private Resort and Spa with 413 reviews

1.7.2 Time and Period of Research

The review used is from the current date of data scraping until the beginning of review started. The time to scrape the review would be done in 23rd of February 2018.

1.8 Writing Systematics

This research is presented into 5 (five) chapters, the systematics are as follows:

CHAPTER I: INTRODUCTION

This chapter briefly presents general explanation of the research. The content includes overview of research object, research background, research questions, objectives, significance of the study, and writing systematics.

CHAPTER II: LITERATURE REVIEW AND SCOPE OF THE RESEARCH

This chapter contains literature review and previous researches which are related to the topic and variables of the research to understand the research.

CHAPTER III: RESEARCH METHODOLOGY

In this chapter, the researcher describes the characteristic of the research, operational and variable measurement, steps, sample, resources, validity, and data analysis method.

CHAPTER IV: ANALYSIS AND DISCUSSION

This chapter provides results of the research systematically to be analyzed by using the methodology and discussion.

CHAPTER V: CONCLUSION AND SUGGESTION

The last chapter of the research presents the conclusion which is derived from all the data processing and analysis and recommendation for further development field of the research.