



**Aalto University**  
School of Business

Mikkeli Campus

**SOCIAL MEDIA COMPETITIVE ANALYSIS AND TEXT MINING**  
A case study in digital marketing in the hospitality industry

Anna Garant

International Business  
Bachelor's Thesis  
Supervisor: Dale Fodness  
Date of approval: 13 April 2017

Aalto University  
School of Business  
Bachelor's Program in International Business  
Mikkeli Campus





**Aalto University**  
School of Business

Mikkeli Campus

**SOCIAL MEDIA COMPETITIVE ANALYSIS AND TEXT MINING**  
A case study in digital marketing in the hospitality industry

Anna Garant

International Business  
Bachelor's Thesis  
Supervisor: Dale Fodness  
Date of approval: 13 April 2017

Aalto University  
School of Business  
Bachelor's Program in International Business  
Mikkeli Campus



**Author:** Anna Garant

Title of thesis: SOCIAL MEDIA COMPETITIVE ANALYSIS AND TEXT MINING: A CASE STUDY IN DIGITAL MARKETING IN THE HOSPITALITY INDUSTRY

**Date:** 13 April 2017

**Degree:** Bachelor of Science in Economics and Business Administration

**Supervisor:** Dale Fodness

### Objectives

The main objectives of this study were to explore the effectiveness of using text mining to analyse the consumer generated content from online hotel reviews. Specifically, this study focuses on demonstrating the helpfulness of such tools in the case of Original Sokos Hotel Vaakuna Helsinki and Scandic Marski in Finland. By analyzing the current trends and patterns of the online reviews of the two hotels, the objective of the study is to understand the extent to which text mining can improve marketing decisions and thus bring value to consumers.

### Summary

The tourism and hospitality industry has changed tremendously due to the emergence of online review platforms such as TripAdvisor.com. This study applies text mining analytics to conduct a content analysis on the social media content provided by hotel guests on these platforms. To gain competitive insights from the data, topic classification and sentiment analysis are used.

### Conclusions

The findings of the research illustrate how topics and related sentiment can be identified from the online content. Although there are several similarities between the data regarding online discussion, the text mining analysis also identified some differences, which have the potential to contribute to gaining competitive intelligence in the industry. Overall, the study illustrates how simple text mining software, which requires little resources from firms can provide beneficial information about the market to hotels in international business.

**Key words:** *digital marketing, hotels, reviews*

**Language:** English

**Grade:**

COVER PAGE

TITLE PAGE

ABSTRACT

TABLE OF CONTENTS

1. INTRODUCTION .....	1
1.1. Background .....	1
1.2. Research Problem .....	2
1.3. Research Question .....	3
1.4. Research Objectives .....	3
1.5. Definitions .....	3
2. LITERATURE REVIEW .....	5
2.1. Introduction .....	5
2.2. Text mining social media .....	6
2.2.1 The process .....	7
2.2.2. Benefits and limitations .....	8
2.3. Text mining in the hotel industry .....	10
2.3.1. Topic classification .....	11
2.3.2. Opinion mining .....	12
2.3.3. Text mining hybrids .....	13
2.4. The role of online reviews in the hotel industry .....	13
2.5. Conceptual framework .....	17
2.6. Conclusion .....	17
3. METHODOLOGY .....	19
3.1. Context of the study .....	19
3.2. Procedures .....	20
4. FINDINGS .....	22
4.1. Phase one .....	22
4.2. Phase two .....	23
4.2.1. Topic classification .....	24
4.2.2. Sentiment .....	27

4.2.3. Demographics .....	30
5. DISCUSSION AND ANALYSIS .....	31
6. CONCLUSIONS .....	35
6.1. Main Findings .....	35
6.2. Implications for International Business .....	36
6.3. Suggestions for Further Research .....	37

## 1. IINTROCTION

### 1.1. Background

Social media has become an integral part of daily communication among and between consumers and businesses (He, Zha & Li, 2013). This transformation in online communication has also changed the traditional view of marketing from a model of one-way communication from sender to receiver into a more dynamic and complex system where the roles of sender and receiver are blurred (Lagrosen & Grundén, 2014). On one hand, customers taking part in dynamic interaction enables businesses to obtain valuable information about customer profiles, preferences, experiences and expectations (Xiang et al., 2017). This is important to note especially in the tourism and hospitality industry, where the emergence of online hotel review platforms is changing the way consumers research their destinations and make decisions in which hotel to stay. Furthermore, sharing recommendations and negative feedback has become a public act rather than individual interaction among consumers and between customers and firms. (Cenni & Goethals, 2017) Therefore, it is increasingly important that firms stay aware of the current trends in online reviews posted by their customers, because the reach of them is wider and more efficient than ever before. One of the most popular of such travel forums is the online review website TripAdvisor.com (Liu et al., 2017), which will be the subject of analysis in this study.

The increasing amount of consumer created content online opens various opportunities for businesses. As the tourism management is a complex process, it is vital that hotels be aware of consumer views to utilize the knowledge in strategic marketing and management (Krajnovic, Bosna & Jasic, 2012). However, the amount of content created by the open communication of social media users and companies is often too large to effectively process through traditional text analysis. Thus, companies can widely benefit from using text mining to process and analyse large masses of unstructured data. Text mining is a technique where computational algorithms are used to discover relevant trends and patterns from textual data. (Yee Liau & Pei Tan, 2014) The methodology has been widely used to analyse internal documents within firms, and extending its capabilities to the field of social media has become a trend only in recent years (He, Zha & Li, 2013). Firms are



beginning to discover the potential of text mining analytics in creating a competitive advantage by screening the environment and doing market research on public data provided directly from consumers themselves. Because of this great potential and recent growth in social media usage in traveling, the tourism and hospitality industry seems to be an ideal field to study social media analytics (Xiang et al., 2017). Therefore, the objective of this paper is to illustrate how text mining tools can be useful in bringing forth relevant trends from these interactive platforms.

## 1.2. Research Problem

In the fast-changing digitalized business environment, companies are constantly generating and collecting large amounts of data. As an example, social media channels serve as platforms for dynamic customer interaction and thus contain large amounts of unstructured qualitative data about customer experiences and perceptions. Especially online travel forums such as TripAdvisor.com, have established a strong position in the travel decision making process of consumers. (Hu, Chen & Chou, 2017) Understanding and using the relevant data provided in these accounts can benefit companies in online marketing and building customer relationships. This can especially benefit the hospitality industry in which the Sokos Hotels and Scandic hotels chains operate, as the industry is facing large changes. With the emergence of new formats of accommodation such as room sharing, traditional hotel businesses are faced with more pressure than before to focus on increasing customer value. Thus, by understanding its own and its competitors' social media trends, hotels have potential to gain a competitive advantage in the industry.

To find the key trends and patterns from the large amount of qualitative data provided in social media accounts, companies can utilize text mining. Through the text mining process, the original text can be categorized into variables to create a semi-structured data base analyzable by the text mining software. By applying different techniques and features, trends, patterns, opinions and relationships between certain topics can be found from the originally unorganized set of reviews. Thus, the text containing qualitative data is turned into a structured data base. The structured data can be visualized and analyzed using text mining tools, and the information can then be utilized to create predictive and descriptive models to improve social media marketing, which this study will aim to illustrate. A problem

in the past has been that this often requires sophisticated technologies and high financial resources. However, with the emergence of new technologies, text mining offers a pool of options for consumer generated content analysis.

### 1.3. Research Question

This study aims at exploring the role of online reviews and text mining in the hospitality industry. Specifically, I would like to know what competitive insights can be derived from text mining hotel reviews from TripAdvisor, regarding Original Sokos Hotel Vaakuna Helsinki and one of its nearest competitors Scandic Marski.

### 1.4. Research Objectives

1. To explore the use and effectiveness of text mining in the hospitality industry.
2. To explore the usefulness of these tools in the case of online reviews regarding Original Sokos Hotel Vaakuna Helsinki and Scandic Marski.
3. To analyse the current trends and patterns of TripAdvisor reviews regarding the above mentioned hotels.
4. To understand the extent to which text mining can improve marketing decisions and increase value to customers.

### 1.5. Definitions

In this section, some of the key terms used in the literature are defined.

Digital marketing refers to promotion, outreach and other traditional aspects of marketing that are implemented using digital technology (Hicks, 2013). Interacting with consumers on social media platforms with the aim of promoting a firm's product or service is recognized as digital marketing in this context.

Text mining is a methodology in which computational algorithms are used to identify relevant trends and patterns in large amounts of existing textual data. Unlike a traditional search engine, text mining helps in identifying relationships between words rather than just finding them in the text. (Clark, 2012)

Unstructured data is term used to illustrate a set of data that is presented in a format that is unorganized and therefore cannot be automatically analysed by machines. Examples include data in the format of text, video and images. (Ganomi & Haider, 2015)

Semi-structured data refers to a set of data with features of structured and unstructured data. In this case, it is used to describe the format of the data base created from the pre-processed text. It has an organized structure but also contains unstructured aspects such as words but can be analysed using automated technology. (Abiteboul, 1997)

## 2. LITERATURE REVIEW

### 2.1. Introduction

Social media has become an integral platform of communication between consumers and businesses (He, Zha & Li, 2013). The emergence of interactive online platforms has also changed the traditional view of marketing from a model of one-way communication to a more dynamic and complex system, where the roles of sender and receiver are blurred (Lagrosen & Grundén, 2014). On one hand, customers taking part in dynamic interaction enables businesses to obtain valuable information about customer profiles, preferences, experiences and expectations (Xiang et al., 2017). The role of sharing experiences and seeking information about destinations has become central in the tourism and hospitality industry, with the emergence of various online travel forums (Chung & Koo, 2015; Jannach, Zenker & Fuchs, 2014). Of such travel forums, TripAdvisor.com has a leading position in terms of quantity of users, destinations and reviews (Hu, Chen, Chou, 2017). However, the amount of content created by the open communication of social media users and companies is often too large to effectively process through traditional means of content analysis. Therefore, companies can widely benefit from using text mining to discover relevant trends and underlying topics in consumer generated social media content. (Liau & Tan, 2014; Qi et al., 2017)

Because of the increasing popularity of sharing experiences through online reviewing platforms, the tourism and hospitality industry seems to be an ideal field to text mining analytics in social media (Xiang et al., 2017). Several studies have already been conducted to apply text mining techniques to the hotel industry (Bjorkelund, Burnett & Norvåg, 2012; Mankad et al., 2016; Guo, Barnes and Jia, 2017; Kim et al., 2017). This literature review will explore the use of text mining in conducting content analysis on social media, especially to gain competitive insights in the hotel industry. To explain the concept of text mining, the literature review will start off with presenting the process of text mining and some of its general benefits and limitations when applied to the social media context. Furthermore, some previous applications of text mining in the hotel industry will be illustrated with case examples from the literature. Because the aim of the paper is fundamentally to explore the

use of the technique in analysing online hotel reviews, the paper will continue to demonstrate the importance of online reviews in the hotel industry. To illustrate the process in a concise manner, this literature review will end in a conceptual framework which depicts the process of text mining consumer generated content with the objective of gaining competitive intelligence.

## 2.2. Text mining social media

As the popularity of using text analytics in the social media context is increasing, the literature provides various examples of how social media text mining is beneficial for companies. Throughout the literature it is identified that the objective for a company to use text mining in the marketing context is fundamentally to gather competitive intelligence from unstructured data to position itself effectively in the marketplace (Leong, Ewing, Pitt, 2004; Gémar & Jiménez-Quintero, 2015). Especially in the hospitality industry, it is increasingly important for organizations to adapt to the rapidly changing environment. Therefore, a company's ability to make strategic decisions strongly depends on their ability to gain knowledge about their own strengths and weaknesses and that of their competitors. In addition, it can contribute to establishing a stronger position in the marketplace. (Bjorkelund, Burnett and Norvåg, 2012; Gémar & Jiménez-Quintero, 2015) Social media content generated by consumers offers a pool of openly accessible and free information for businesses to exploit for various operational improvements in the marketing sector, for instance. Harrysson, Metayer and Sarrazin (2012) state that because of the changes in the environment, companies need to develop new and effective ways by which to communicate and interact with consumers and collect information. When there is too much data for manual qualitative analysis, text mining offers an effective solution for finding the most relevant trends within the content. The following section will explain the basic concept of automated text analytics, which will be followed by an exploration of some of the benefits of using such techniques in business.

### 2.2.1 The process

Before going further with how text mining can assist hotels in competitive positioning, a brief overview of the concept is in place. Text mining is an automated technique that uses computational algorithms to extract meanings and patterns from already existing text (Lau, Lee & Ho, 2005; Clark, 2012; He, Zha & Li, 2013; Gémar & Jiménez-Quintero, 2015; Qi et al., 2017). As stated by Xiang et al. (2017), primary research has traditionally been conducted by communicational studies such as surveys and interviews designed to collect data directly from consumers. Text mining however, allows for similar analysis by exploiting existing information online. Thus, the method does not necessarily discover new knowledge, but rather helps in analyzing and identifying the relevant information from large amounts of currently existing unstructured data. In addition, text mining aims to identify relationships between words in sentences rather than just finding words in the way of a search engine. (Clark, 2012; Upshall, 2014) There are many different techniques used under the general term of text analytics, which include text summarization, clustering, string matching, classification and sentiment analysis (Leong, Ewing, & Pitt, 2004; He, Zha, Li, 2013; Upshall, 2014).

The literature explores many different techniques of text mining applications, which include different procedures in the analysis process. The basic concept of text mining involves reorganizing the natural language from its unstructured format into a semi-structured data base which can be analyzed by computer software (Kotu & Deshpande, 2014). Gandomi and Haider (2015) define the term unstructured as data which is presented in an unorganized format due to which it cannot be analyzed by machines. The authors mention text, video and images as examples of this. Defining semi-structured data appears to be reasonably complex due to the various formats in which it can appear. However, it can be characterized to have a structure that is analyzable by automated computational techniques and still have elements of unstructured data, for example, unstructured text (Abiteboul, 1997).

Generally, the text mining process can be divided into four stages, which include defining the concepts and the context for mining, collecting data, dictionary construction and ultimately analysis and visualization. (Gémar & Jiménez-Quintero, 2015) The unstructured collection of text especially in the social media context often requires some preprocessing

to make the results more accurate and useful. This step includes changing text into lowercase, deleting words that are meaningless in the context, and that occur too frequently or too seldom, and deleting suffixes to detect only the main body of a word if it occurs in a conjugated form. (Mankad et al., 2016) After performing these steps to bring structure to the content, there are various predictive and descriptive functions that can be applied to the data, including sentiment analysis, clustering and classification (Kotu & Deshpande, 2014). There are several tools available that can perform these functions. An example of such mining software is RapidMiner, which offers the basic technology to conduct text analysis for social media content such as tweets and online reviews (Kotu & Deshpande, 2014). This tool will be used in the case study presented further in this paper.

### 2.2.2. Benefits and limitations

As stated earlier, text mining can be applied to process and analyze large amounts of quantitative data. Furthermore, it can make gaining insights from consumer generated content more efficient in addition to providing several other benefits for managerial decision making and marketing operations. (Harrysson, Metayer & Sarrazin, 2012; Clark, 2012; Gémar & Jiménez-Quintero, 2015) This section will explore some of the ways in which text mining can assist in businesses in environmental screening and market research, in addition to identifying some areas of improvement for the technology in its current state.

The literature offers insights into to the various benefits of text mining analytics in the business context. For instance, Harrysson, Metayer, & Sarrazin, (2012) propose that text analytics, especially sentiment analysis, can be used to estimate market size and customer opinions about products or services. The authors state that traditionally such research has been conducted through focus groups and field observation about consumer behavior, which can be both ineffective and costly. This suggests that text mining techniques can reduce marketing research costs significantly when used to map consumer demographics and opinions. In addition, Cesaroni and Consoli (2015) indicate that the main reason why social media offers a variety of new opportunities for businesses is that little resources or technological expertise is required to utilize these tools. On the other hand, it was earlier argued that in its current state text mining involves high costs related to online

infrastructures, technology customization, transactions, and hiring experts to operate sophisticated technology (Linoff & Berry, 2011; McDonald & Kelly, 2012). However, as social media is an openly accessible source for data the transaction and infrastructure costs are adequately insignificant contributors to cost in this context. The authors state that although there are some tools available for basic text mining procedures, customizing technology for high-end research is costly in addition to the staffing costs for experts in the field. Nonetheless, social media provides an underexploited source for deducing consumer insights, which can be done without expensive and sophisticated technology.

According to Liao and Tan (2014), text mining enables content analysis in a quicker and more efficient way, which can be used as an advantage especially for customer oriented companies such as in the hospitality industry. The authors further argue that firms can use this to predict upcoming trends, analyze the demographics of customers and track their competitors in the same categories. Ultimately, text mining analysis can enable businesses to make informed decisions based on competitive intelligence and take proactive steps to gain a competitive advantage. Furthermore, Uppshall (2014) alleges that automated text mining techniques are significantly more accurate than manual text analysis. In addition, Liao and Tan (2014) suggest that text mining results can be interpreted as less biased than survey results because the data used is extracted directly from the customer interactions on social media. The authors suggest that as the online content is generated voluntarily by consumers, the opinions expressed can be seen as sincerer and more direct. Therefore, it appears that in some cases mining social media content can, in fact, be more efficient and provide more accurate results than traditional means of market research.

However, there literature suggests that the current technologies still contain some constraining features. Lau, Lee and Ho (2005) declare that regionally tailored dictionaries are a major issue especially with cross cultural studies. This challenge can especially be encountered on social media where slang words are commonly used and spoken language is written. Furthermore, Nguyen, Shirai and Velcin (2015) argue that sentiment analysis on social media is especially complex due to the short text that contain misspellings and grammatical errors. In addition, Uppshall (2014) has earlier argued that executing corporation level text mining requires experts with comprehensive knowledge of the software, which can be a challenge for many businesses. Liao and Tan (2014) also add that the techniques of sentiment analysis fail to identify sarcasm, which is a major weakness



when studying social media. Therefore, despite the efficiency and new possibilities that text mining can offer businesses, there are still some hindering factors especially in the social media context.

Text analytics can also be used to mine sentiment from social media content to forecast future market trends. In the finance sector the effects of social media sentiment mining on stock market prediction have been widely studied (Khadjeh Nassirtoussi et al., 2014; Nguyen, Shirai & Velcin, 2015; Sun, Lachanski and Fabozzi, 2016). Ngyen, Shirai and Velcin (2015), for instance, present a method by which stock prices can be predicted using sentiment and opinion mining. The study concluded that results derived through the text mining method were, in fact, more accurate than those found with other previous techniques. This indicates that text mining and especially sentiment analysis can be used to predict future trends based on online consumer behavior. This can also be applied in the digital marketing sector to enhance consumer understanding and target products and marketing campaigns accordingly (Sun et al., 2015). However, at least in the stock market context Khadjeh Nassirtoussi et al., (2014) argue that even experts in the field cannot draw a correlation between opinions expressed online and stock price changes with certainty. The maturity of the literature, however, indicates that there are predictive capabilities to some extent in text mining analysis. The extent to which these capabilities provide accurate results, and whether they can be applied to the marketing context offer areas for possible future research.

### 2.3. Text mining in the hotel industry

Gaining knowledge about customer satisfaction has been a challenging but central issue in the hospitality industry for years. Traditionally, companies have approached this issue by collecting feedback through different means, including surveys, mystery shoppers and comment inquiries. These approaches have been identified as problematic because of the high volume of unstructured textual data, which is difficult to analyse in an effective and reliable way, compared to quantitative data analysis. (Mankad et al., 2016) Another issue with traditional hotel guest feedback methods is the insufficiency of uncompleted feedback forms and the reluctance to disturb hotel customers during their stay (Qi et al., 2017). Since

the emergence of text mining technologies, various studies have been conducted to apply such tools to different aspects of the hospitality industry. There are several studies that conclude that using text mining tools to extract competitive intelligence from social media can, in fact, improve the financial performance, competitive positioning and marketing activities of a hotel company. (Gémar & Jiménez-Quintero 2015; Xu & Li 2016; Guo, Barnes & Jia, 2017) The following sections will present some of the commonly used methods of text mining and some cases from the hotel and tourism industry.

### 2.3.1. Topic classification

One commonly used text mining approach throughout the literature is mining common underlying topics from the textual data using Latent Dirichlet Allocation (LDA) (Mankad et al., 2016; Guo, Barnes and Jia, 2017). To demonstrate topic classification in the hotel industry, Mankad et al. (2016) used the LDA method to study online hotel reviews. The authors encountered five topics in the study, including amenities, location, transactions, value, and experience. According to the authors, classifying topics using LDA can assist managers in making strategic decisions. For instance, by identifying the major discussion points in the collected feedback online, hotels can have more information about what aspects suggested in guest reviews to improve. To add to these findings, Guo, Barnes and Jia (2017) in larger study that the most influential topics in hotel reviews were room experience and service quality, which could further be divided into various subcategories. Xu and Li (2016), on the other hand, offer insights on the usefulness of Latent Semantic Analysis, which is another similar text mining technique to identify underlying meanings and topics from a large collection of writing. This technique was used to identify which attributes affect consumer satisfaction in different types of hotels distinctively from both negative and positive reviews. By using this technique on a random sample of 3480 reviews, the study found distinct themes about both positive and negative associations which varied depending on the type of hotel. Subsequently, this shows how online reviews provide a valuable source of knowledge on the base of which managers can make strategic decisions. By categorizing the topics underlying in a large content of qualitative data, it is easier for managers to assess the major issues and discussion points that consumers present in online platforms. However, the LDA technique is currently a reasonably intensive

and resource consuming approach to text mining, which is why it ineffective to be adopted by hotel companies (Guo, Barnes & Jia, 2017). The literature seems to suggest that because of the service oriented nature of the hospitality industry, identifying and categorizing major discussion points and topics can lead to more effective decision making.

### 2.3.2. Opinion mining

Opinion mining, also referred to as sentiment analysis, is also commonly used to discover sentiment from customer feedback, such as online hotel reviews. Bjorkelund, Burnett and Norvåg (2012) provide an example on mining opinions and sentiment from online hotel reviews on major travel websites. The authors then use the results to map favorable and unfavorable destinations using the Google maps application. By creating this prototype, the authors suggest that the application could help consumers evaluate their destination as well as managers analyze their market position. However, the literature identifies a number of limitations related to current sentiment analysis techniques (Bjorkelund, Burnett & Norvåg, 2012; Kim et al. 2017). For instance, according to Bjorkelund, Burnett and Norvåg (2012), some comments contain both negative and positive sentiment, which makes it difficult to categorize as either. In addition, some reviews are merely descriptive and therefore cannot contribute to such an analysis. Another issue with sentiment analysis, as described by Ryu and Lee (2016), is that the sentiment related to some words are highly dependent on the context in which they are used. The author provides an example of the word “unpredictable” which can be positive when describing a movie plot but negative when explaining the functionality of a product, for instance. Furthermore, Kim et al. (2017) disclose that although positive and negative perceptions can be extracted from the content, the technology fails to explain reasons behind these feelings. The literature implies that opinion mining can be used to discover consumer preferences and opinions about different aspects of hotel service, but results should be examined with caution due to the above-mentioned limitations. Often, the method provides an overview of the overall sentiment present in the set of data, however, the results should not be generalized or taken as definite.

### 2.3.3. Text mining hybrids

To reach more comprehensive results, studies often combine many different text mining techniques. The literature illustrates that to reach findings that provide the most benefit to companies through content analysis, it is often not enough to apply only one feature of text mining. To illustrate the use of text mining hybrids, G mar and Jim nez-Quintero (2015) present a study which determines the effect of social media presence and customer comments on the return on equity of hotels. The methodology applied in the research included sentiment, passion and reach analysis from the most popular social media sights such as Facebook, Twitter, and YouTube. The study shows that return on equity was effected by low ratings in passion, which is defined as the likelihood that comments originate from the same author versus that each comment is submitted by different authors. Thus, the research found that the larger the number of commenters was on social media, the greater the influence is on the profitability of the hotel. Kim et al. (2017) provide more insight on how sentiment analysis can be used to mine consumer opinions from online hotel reviews but also adds co-occurrence analysis to reach more comprehensive results. In this comparatively large scale study, the authors discovered how customers perceived different aspects of the hotels services. More specifically, 14 different aspects of the hotel reviews were defined to be assigned values, which then determined whether the overall sentiment towards them was positive, neutral or negative. Thus, more detailed findings could be made from the data when compared to traditional sentiment analysis. Overall, it appears that by combining different text mining methods together businesses can, indeed, reach more comprehensive results based on which to make more informed decisions.

### 2.4. The role of online reviews in the hotel industry

After exploring the literature on text analytics as a tool for content analysis, and how it has been utilized to analyse the social media channels and different aspects of the hotel industry, it seems appropriate to discuss the literature on why the technique is useful specifically in the context of online reviews. As discovered earlier in this literature review, the tourism and hospitality industry has been profoundly changed by the advances in internet technology and social media applications. This quite recent development has made

online tourism applications the primary channel for information search regarding travelling decisions (Hu, Chen & Chou, 2017).

A survey conducted by Ady and Quadri-Felitti (2015) reveals that approximately 95% of travelers read reviews before booking a hotel. Furthermore, the authors argue that over a third perceive these reviews to be the most important contributor in their decision. This proves that the content available in the social media accounts has a notable impact on the travelers' decisions. In addition, this supports the claim that managers should be aware of the major trends in customers' online reviews because they do, indeed, contribute to other consumers' purchasing behavior. This illustrates the importance of the consumer generated content in online travel forums, and the fundamental purpose behind finding new and effective ways for content analysis. This last section of the literature review aims at exploring the role of online reviews in the hospitality industry, to illustrate why travel websites such as TripAdvisor are optimal targets for automated text analysis.

Traditionally research on consumer satisfaction has largely been derived from quantitative data available from numerical online ratings. However, the emergence of text mining technology has enabled more detailed research into consumer minds through the analysis of qualitative online reviews. (Hu, Koh and Reddy, 2013; Guo, Y., Barnes & Jia, 2017) Hu, Koh and Reddy (2013) propose that although a correlation between online numerical ratings and sales cannot be identified, there is a straight correlation between sentiment in online reviews and sales. Therefore, it is reasonable to suggest that textual reviews contribute more to consumers purchase decisions and thus are of more value to firms. Although they may have a positive impact on destination image, numerical ratings should not be relied on too much, as they can differ quite notably from the related text. A practical example of this will be presented later in this paper.

The power of online reviews is constantly growing as communication is increasingly moving towards online formats. Furthermore, as opinions are widely shared through online reviews, they can be seen as a form of electronic word of mouth. When compared to traditional face-to-face word of mouth, the reach and speed of electronic word-of-mouth is much greater, which makes it a more powerful source to shape consumer's opinions and perceptions of the service. (Xun & Li, 2016) This enables both the consumers and the companies to benefit in terms of efficient communication and easy access to information. Furthermore,

the low barriers to spread online word-of-mouth generates more content from which businesses can derive competitive insights. For consumers, this allows access to a wider pool of information to shape opinions about potential destinations (Agag & El-Masry, 2016). According to a study by Abubakar and Ilkan (2016) conducted in the medical tourism sector, positive electronic word-of-mouth was found to have a positive influence also on travel intentions and trust towards the destination. Therefore, to build trusting relationships with customers, managers should understand current trends of their own and their competitors' social media accounts.

Based on the current evidence on the matter, it seems fair to suggest that companies should take proactive steps to encourage positive word-of-mouth in their social media channels as well as third party travel forums. According to Amaro and Duarte (2015), the main factors that affect customers' online purchase intentions for hotels online are attitude towards online purchasing in general, perceived relative advantages and perceived risk in addition to the complexity and communicability of the online platform. Agag and El-Masry (2016) argue that the same factors influence interaction and electronic word-of-mouth in an online travel community such as TripAdvisor. Furthermore, the study indicates that consumers are more likely to promote a product or service in such online communities if they are perceived to add more value than those of competitors. Therefore, as stated by the authors, consumers that participate in the online travel communities are more likely to generate loyalty to the firm and spread positive word-of-mouth. Through text mining online content, marketing managers can remain informed of the dominant trends in these travel communities, not only to enhance their product or service, but also to provide an attractive and consumer friendly ambiance in the online community. Therefore, it is important to be active in replying to comments and reviews in addition to being aware of the macro trends within the mass of reviews.

For marketing managers, there are also several positive findings about the importance of online reviews. As demonstrated by Pauwelsa, Aksehirlia and Lackman (2016), once online discussion about the brand is started it is likely to continue among consumers without additional effort from the company. However, Arroyabe (2016) argues that the main contributors to dynamic customer interaction are the company's digital marketing investment and the actions of the person in charge of creating strong relationships with followers. The author further argues that these aspects are also key to generating customer

loyalty and building and attractive brand image on social media platforms. This indicates that the effort to maintain discussion in social media communities mainly originates from a company's actions. According to these findings, companies should focus resources on maintaining an active position not only in its own social media channels, but also remain visible on travel forums. Agag and El-Masry (2016) add that it is, in fact, by positive word-of-mouth that companies can reduce customer service costs and build deeper relationships with customers, thus generating more demand for services. Furthermore, online content generated by consumers is easily available, accessible regardless of location and time constraints. Through personal profiles it can offer significantly more information on customer segments and demographics than any traditional way of data collection for such purposes (Kim et al., 2017; Guo, Barnes and Jia, 2017). Therefore, it seems fair to say that online reviews on third party websites offer a low-cost and wide pool of resources that businesses should utilize to their full potential.

On the other hand, as stated by Xu and Li (2016), the ranking criteria in such online reviews also varies greatly depending on the expectations, perceptions and desires of each individual. In addition, consumers' interpretations of online word-of-mouth are different, which makes measuring the actual effects of the content reasonably complex. However, through the use of text mining managers are able to gain an overall view of the content generated by customer reviews, which can be used to create a competitive advantage (Liau & Tan 2014). According to Reimer and Bankenstein (2016) the issue of whether consumers perceive other reviews as trustworthy or manipulative is also a contributing factor to purchase intentions. The authors add that this is especially challenging when comments are posted anonymously. However, many of the mainstream social media channels provide sufficient profile identification features to mitigate this issue.

As illustrated by the literature, the emergence of text mining analytics and increasing popularity of online reviews on travel forums offer a wide variety of opportunities for digital marketing. The following conceptual framework summarizes the process of utilizing text analytics on consumer generated content to build competitive intelligence.

## 2.5. Conceptual framework

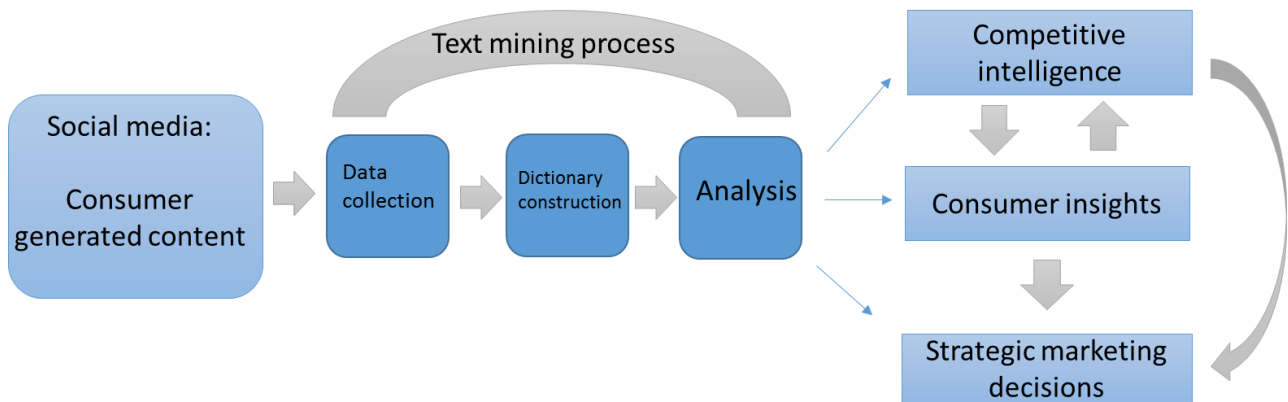


Figure 1: Social media text mining process

## 2.6. Conclusion

The purpose of this literature review was to explore the benefits, limitations and prior studies of text mining in the social media context, with a specific focus on the hospitality industry. In addition, the literature offered insights to why online hotel reviews are an optimal object for text mining analysis. Figure 1 provides a conceptual framework which illustrates the process of mining consumer generated content to obtain valuable information about consumers in order to make more informed strategic marketing decisions. As explained previously in this literature review, social media contains large amounts of data provided directly by customers, and is therefore an optimal object for text mining. Figure 1 demonstrates the key steps for the text mining process which included collecting the data, constructing the dictionary and applying different analytical features to gain sufficient results. This can enable businesses to gain valuable insights about consumers, which contributes to building competitive intelligence in the market. Ultimately companies can utilize these insights to make strategic decisions to strengthen their competitive position, which is demonstrated in the conceptual framework.



Based on the literature, several advantages related to the technology were identified, including collecting customer knowledge in a cheap and efficient way and consequently gaining competitive intelligence. These insights can ultimately assist managers in making informed operational decisions in terms of digital marketing in the hospitality industry, for instance. However, some limitations were also identified, such as literary restrictions and the complex and sophisticated nature of some of the current technologies. With the expanding amount of consumer generated online data, hotels have access to a pool of information about consumer perceptions of their business and that of their competitors. This creates an optimal environment for utilizing text mining tools, which can simultaneously lead to lower costs in terms of market research. Overall, text mining remains an underexploited tool for social media content analysis, and could be further developed to assist firms in daily operations.

### 3. METHODOLOGY

#### 3.1. Context of the study

As discussed in the literature review, the role of online customer reviews has grown notably in recent years, which especially affects the hotel industry. Furthermore, these online reviews have been proven to impact the hotel choices of travellers. Therefore, it is vital for hotel companies to not only be aware of the discussion evolving around their services but also screen their competitors. To illustrate how text mining can be used in analysing consumer generated content, two hotels were chosen for the subject of the competitive analysis. The basis of choosing these hotels for the analysis were their close location and similar backgrounds. Both hotels are part of large hotel chains in Finland and other Nordic countries, with guests from many countries.

The platform from which reviews were chosen to be extracted was one of the most popular online travel review forums, TripAdvisor. The platform that enables consumers to rate and review destinations was founded in 2000, and its headquarters are located in the United States (Vásquez, 2011). According to Fuggle (2017), over 6.2 million firms use TripAdvisor in more than 128 000 destinations, and every minute over 200 new contributions are made by users on the website. Therefore, the sample data was extracted from a website in which user contributions truly have potential to influence the success of hotels. Furthermore, the author argues that 70% of consumers check approximately 200 online reviews before booking a hotel, which indicates the importance of the consumer generated content in hotel booking decisions.

In order to effectively demonstrate how a competitive analysis can be done on online reviews on TripAdvisor, two competitive hotels with active guests on the online platform were chosen. To get a diverse sample of reviews in the English language, the location of the subject hotels was chosen to be the Helsinki city centre. The first subject hotel is Original Sokos Hotel Vaakuna from the Sokos Hotels chain, and the second is Scandic Marski from the Scandic Hotels Group. The Scandic Hotels Group is the largest hotel chain in the Nordics, with 230 destinations in 7 countries (scandichotels.com). Despite its

Swedish ownership, Scandic Hotels have also established a strong position in the Finnish market, with 24 hotels in 16 Finnish Cities. Sokos Hotels, on the other hand is owned by the Finnish S Group, and has over 50 hotels in Finland in addition to some in St. Petersburg, Russia and Tallinn, Estonia (sokoshotels.fi). In addition, both hotels were recommended in the top ten selection of hotels near the Helsinki city centre train station. In this report, Original Sokos Hotel Vaakuna Helsinki will be referred to as Sokos and Scandic Marski as Scandic.

### 3.2. Procedures

To answer the research question of what competitive insights can be derived from online reviews concerning Sokos its competitor Scandic, a competitive analysis was conducted in three stages. These stages included pre/processing the text extracted from the online reviews, applying text mining techniques, and finally analysing the results for competitive insights. The aim of the analysis is to demonstrate how relevant trends and patterns can be derived without great financial investments, data analysis experts, or sophisticated technologies. The tool used for the text mining analysis is called RapidMiner, and is available for downloading online. The tool enables simple but powerful text mining procedures with different automated functions and no need for coding, which will be explained in more detail in the following section. The data from TripAdvisor was extracted and saved into an excel sheet from which it was further processed using the software.

Before moving to the text mining process, some quantitative data regarding the numerical ratings and number of reviews on TripAdvisor for both hotels was gathered manually and saved to an excel sheet for further analysis. This information included the numerical ratings of both hotels and the response rates to feedback by official hotel representatives on the TripAdvisor website. For the analysis, reviews in English language were collected from the six most recent months from the analysis, September-February. The total number of reviews extracted from the website from this time period was 206, of which 111 referred to Sokos and 95 referred to Scandic.

To further analyse the social media content, text mining techniques were applied to analyse the unstructured text from consumer reviews. The text mining process itself can be divided into multiple steps. To create a semi-structured data base for analysis, the unstructured data was pre-processed using a variety of features on the analysis software. Firstly, the text was tokenized using non-letter characters as breaking points, meaning that each review was split into individual words. Following this procedure, all of the letters were transformed into lower case, and English stop words were removed from the word list. Stop words include words that occur in the text often but are irrelevant in the text mining process. Examples of these include words such as “and”, “then” and “that”. Next, n-Grams were generated from the tokens. The definition of n-Gram is a sequence of tokens that occur in the text in the length of n. In this case, n was attributed to 3 to find series of consecutive words in the maximum length of three. This step demonstrates which words are likely to occur together in the text, bringing more insight to the data than a basic word count.

Next, the data was analyzed in terms of sentiment. The sentiment analysis, also referred to opinion mining was conducted using the aspect based sentiment analysis feature available in RapidMiner studio. The process assigns different values to the pre-assigned aspects, which it then identifies from the data set created from above. Subsequently, the program creates a table illustrating which of the 15 different categories assigned to hotel reviews were identified to be positive, neutral and negative. In addition, the feature displays from which document the sentiment was identified, illustrating the differences in opinion for the two hotels. The end result is a semi-structured database illustrating each pre-assigned aspect and the related sentiment found from the data.

## 4. FINDINGS

### 4.1. Phase one

In phase one of the analysis, quantitative data from the TripAdvisor website was analysed to form a basic understanding of the online position of each hotel. In total, the TripAdvisor page for Sokos contained 1,705 reviews at the end of February 2017, of which 933 (54.7%) were written in English. For Scandic the total number of reviews was 1279, of which 684 (53.5%) were in English. As the TripAdvisor website displays a view of how the ratings of each hotel is distributed it provides quantitative data for instant comparison. Table 1 presents an overview of how ratings from one to five stars were distributed between the two hotels. As seen in the table, the average of ratings was higher for Sokos when compared to those of Scandic. Sokos gained higher volumes of excellent and very good ratings, whereas Scandic had a greater emphasis on average ratings.

**Table 1: Distribution of total English ratings**

Traveler rating	Sokos		Scandic	
	Number	Percentage	Number	Percentage
Total (English)	933		684	
(5) Excellent	235	25.19	138	20.18
(4) Very good	473	50.70	313	45.76
(3) Average	177	18.97	167	24.42
(2) Poor	38	4.07	49	7.16
(1) Terrible	10	1.07	17	2.49
Average of ratings	3.95		3.74	

For the text mining analysis, reviews were collected from the time frame of six months from September 2016 to February 2017. Within this time frame, the number of English reviews for Sokos was 111, whereas for Scandic it was 95. Table 2 illustrates the distribution of ratings for the sample that was collected to be the object for the text mining analysis. As seen in the table, the distribution of the sample is quite similar to that of the entire

population. However, the average of ratings is slightly lower for both hotels in the sample period.

**Table 2: Distribution of sample English ratings**

Traveler rating	Sokos		Scandic	
	Number	Percentage	Number	Percentage
Total (English)	111		95	
(5) Excellent	27	24.32	19	20
(4) Very good	55	49.55	38	40
(3) Average	20	18.02	28	29.47
(2) Poor	7	6.31	4	4.21
(1) Terrible	2	1.80	6	6.32
Average of ratings	3.88		3.63	

Furthermore, data regarding the response rate was collected for both hotels. The frequency of response was calculated by dividing number of official responses with the total number of reviews. At the end of the sample period, a representative from Scandic had replied to every English review on their page regardless of whether it was positive or negative. Therefore, the calculated response rate for Scandic was a full 100%. During the sample time period Sokos, on the other hand, had a slightly lower rate of response of 93,7%. The difference in responsiveness was caused by Sokos not replying to some of the latest posts on its page, which seems to suggest that they could improve the speed of responding to consumer feedback. Despite the lack of timely responses, both hotels had official representatives reply to every review posted in systematic and comprehensive manner.

#### 4.2. Phase two

In phase two, text mining procedures were applied to the pre-processed text that was extracted from tripadvisor.com. RapidMiner allows several different options for analysing and visualizing the text, the most relevant of which are illustrated in this section of the study. Firstly, common discussion points were identified using the N-Gram feature. This was followed by an aspect based sentiment analysis, which aimed at revealing opinions

and feelings related to specific aspects typical to hotel reviews. In addition, some statistics about the demographics of the reviewers were collected using basic text mining procedures. This section will go through the findings of the text mining analysis each feature at a time, starting from topic classification and continuing on to the results of the aspect based sentiment analysis.

#### 4.2.1. Topic classification

In order to discover connections between commonly used phrases and words within the text, n-Grams with the maximum length of three tokens were generated. This feature was used to create a database which displays the total occurrences of common words and word sequences in both documents. In addition, the data base shows the division of the number of occurrences in the data set of Sokos and Scandic individually. Multiple clear discussion points can be recognized from the data. The key findings are summarized below.

##### (1) Location

Perhaps the most visible topic for both hotels was the central location. The data shows that expressions indicating positive associations with the central location were expressed multiple times regarding both hotels. For Scandic, there were 110 such occurrences, and for Sokos the number of occurrences was 82. Table 3 illustrates the word sequences found regarding location. In addition, the table displays the total occurrences of the pattern, if the attribute was found in both documents, and classifies the number of occurrences in for both hotels. A similar table was created for all of the attributes and major topics identified from the data. The commonality of location as a discussion point for the hotels is not surprising as location was one of the elements on which the decision of sample hotels was based. Therefore, this finding, although central, does not contribute much to the study. However, the results show that although Scandic had less reviews in the set period from September to February, guests have been more active in expressing their satisfaction with the location when compared to Sokos.

**Table 3: N-Grams for location**

Word	Attribute name	Total occurrences	Document occurrences	Scandic	Sokos
great_location	great_location	24	2	19	5
good_location	good_location	23	2	14	9
city_center	city_center	20	2	13	7
excellent_location	excellent_location	20	2	7	13
walking_distance	walking_distance	15	2	8	7
perfect_location	perfect_location	12	2	8	4
location_good	location_good	11	2	3	8
center_helsinki	center_helsinki	10	2	3	7
location_city	location_city	9	2	7	2
location_excellent	location_excellent	9	2	5	4
central_location	central_location	8	2	4	4
middle_city	middle_city	8	2	4	4
location_city_center	location_city_center	6	2	5	1
location_nice	location_nice	6	2	5	1
nice_location	nice_location	6	2	3	3
location_perfect	location_perfect	5	2	2	3
Total		192		110	82

## (2) Service quality

Another common discussion point found from the reviews was the quality of service. The list of n-Grams included multiple notions of good service and friendly staff, with a total of 58 occurrences. These included word sequences such as “good service” and “friendly staff”. Considering all the synonyms and different expressions indicating good service quality, Scandic appeared to have a leading position with 33 in-text occurrences whereas the number for Sokos was 25. This supports the claim that for some reason the guests of Scandic have been more verbal in indicating their satisfaction with the staff and service than those of Sokos in the given period.



### (3) Breakfast and restaurant

In this category, there appeared to be many mentions of the hotel breakfast and restaurant linked with positively charged notions in the data. There was a total of 58 occurrences of different word sequences indicating positive experiences with the hotel breakfast and restaurant. Unlike in the previous categories, the distribution of occurrences for the two hotels was quite balanced with 29 for Scandic and 28 for Sokos. These findings indicate that reviewers were generally content with the restaurant services for both hotels and express it equally in their feedback online.

### (4) General Satisfaction

The text mining process identified several indications of general satisfaction towards the hotel. In addition, word sequences such as “recommend hotel” or “I recommend” occurred 13 times within the data set. The number total occurrences of expressions such as nice, good or excellent hotel was 61. Of these mentions, the majority were once again targeted towards Scandic, with the total of 35. Thus, the number of indications of general satisfaction for Sokos was only 26. However, when compared to the numerical ratings expressing satisfaction, the results are quite contradictory, as Sokos scored higher in the overall ratings in addition to the average ratings within the six-month period. This seems to suggest that the guests of Sokos have expressed their opinions with an emphasis on the numerical rating and less expressing these experiences in the reviews.

### (5) Rooms

One of the topics apparent in the data with the clearest division between hotels was satisfaction towards different attributes in the hotel rooms. There were notions of the rooms as being good, excellent, superior, and clean. In addition, the text mining process identified word sequences relating to features of the hotel rooms, for instance, the words “comfortable beds” occurred together four times in the Sokos document and only once in that of Scandic. However, the division of room satisfaction was greatly lenient towards Scandic, as the number of occurrences for positive room associations was 32 for Scandic out of the total number of 49.

### (6) Renovation

The only clear topic, which was found in only the other data set was indications for the need for renovation. There was a total of 69 notions found in the set of reviews that referred to

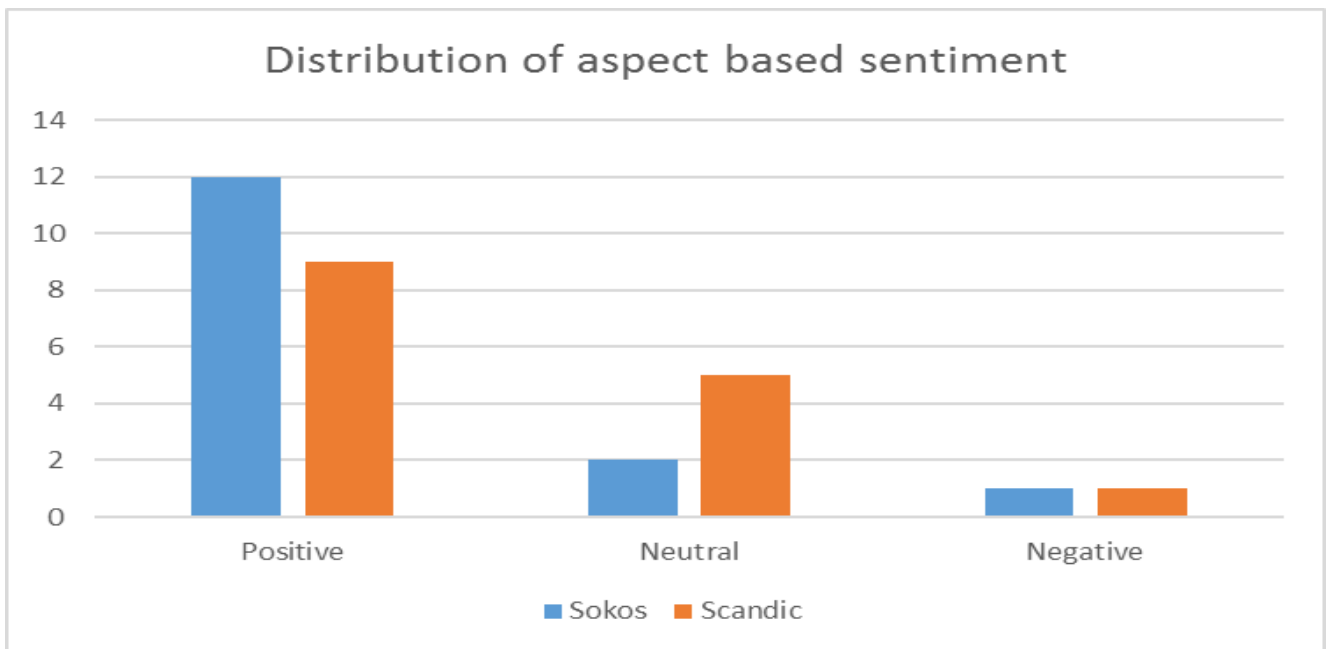
this issue, of which no less than 47 were aimed at Scandic. This indicates that despite the dominant number of occurrences in all of the previous positive topics, Scandic was the only hotel with clear trends for improvement recommendations. Thus, although the guests of Scandic seemed to be more verbal about the positive associations in their online reviews, they were also more active at expressing needs for improvement. The most common topics and trends in the data collected for Sokos, however, did not have any clear indications for improvement.

Overall, by creating a simple data base to illustrate word sequences and the number of occurrences of each, several trends could be identified from the mass of reviews. Although some sentiment involved in the discussion could be identified from the word sequences with positive words, the content could be mined for more in-depth information about the common opinions present in the reviews. Therefore, a sentiment analysis was conducted on the data, which will be further explained in the following section.

#### 4.2.2. Sentiment

Another commonly used text mining approach especially in a social media context is sentiment analysis. Sentiment refers to attitudes, thoughts and feelings, which can be identified from consumer generated content using specific text mining procedures. In this study, a more detailed approach called aspect based sentiment analysis was used to identify the opinions related to different parts of the hotels' performance. An analysis for aspect based sentiment usually requires two phases: extracting aspects and assigning sentiment. RapidMiner studio provides a feature directly aimed at analyzing hotel reviews. The feature is called aspect based sentiment analysis, and it analyzes each of the aspects of a certain product or service and assigns values to indicate whether they are positive, negative or neutral. The RapidMiner tool allows users to choose from a few pre-assigned aspect models to make it easier to analyze aspect based sentiment from reviews for cars, hotels, airlines and restaurants. The categories in which the general sentiment was analyzed, which were pre-defined by the miner were the following 15: view, comfort, food/drinks, room amenities, cleanliness, Wi-Fi, customer service, beds, payment, location, value, staff, design, facilities, and quietness.

Figure 1 illustrates the general distribution of the aspect based sentiment between the two hotels. As can be seen from the figure, the majority of opinions in all of the aspect categories were positive for both hotels. The number of neutral aspect categories varies for both hotels, with the score being only two for Sokos and five for Scandic. Furthermore, negative sentiment was detected from only one category for both Sokos and Scandic. When looking at the figure in more detail, it can be seen that Sokos performed better in terms of having mostly positive sentiment detected for more aspects than Scandic. Subsequently, the findings suggest that reviews towards Scandic were neutral in more categories than those targeted towards Sokos.



**Figure 2: Distribution of aspect based sentiment**

To explore the general opinions of the hotel reviews even further, aspect based sentiment can be analyzed within the different aspect categories, as explained earlier in this section. Table 4 and Table 5 demonstrate the distribution of aspect based sentiment within the 15 assigned categories. As seen in table 4, the Scandic data set contained 5 categories in which the sentiment remained neutral, which include room amenities, cleanliness, beds, value and design. When compared to the findings for Sokos, which are shown in Table 5, there are only two categories which were classified as neutral. These aspects were customer service and facilities. Thus, the emphasis on positive associations is quite different for the subject hotels although they contain many similar aspects as well.

However, the negative associations were limited to one category for both hotels – pricing. To illustrate what kind of comments lead to the negative categorization, the data set was searched for examples of comments with negative sentiment towards price. For Scandic, some examples of comments about pricing are “The price was way too high for the lack of quality” and “The value is just ok but prices are very high”. Similarly, some comments containing the mention of price for Sokos included notions like “Very nice hotel but a bit over priced in my opinion” and “Maybe the quality of the room did not match the price”. However, it should be noted that the reviews also contained positive opinions about the price. The tool labels the aspects as positive, neutral or negative based on the majority of sentiment it finds in the data, which leads to a generalization in terms of which value is assigned to each aspect. This applies to all of the aspect based sentiment found in the analysis.

**Table 4 Aspect based sentiment for Scandic**

Scandic		
Positive	Neutral	Negative
View Comfort Food/Drinks Wi-fi Customer service Location Staff Facilities Quietness	Room amenities Cleanliness Beds Value Design	Payment

Sokos		
Positive	Neutral	Negative
View Comfort Food/Drinks Room amenities Cleanliness Wi-fi Beds Location Value Staff Design Quietness	Customer service Facilities	Payment

**Table 5 Aspect based sentiment for Sokos**

### 4.2.3. Demographics

Another interesting finding were the diverse demographics of the users providing reviews on the subject hotels' pages on TripAdvisor. The division of home countries of the guests providing feedback is considerably different for both hotels. By creating an n-gram data base of the countries of origin of the reviewers, it can be seen that the guest network interacting on the Scandic TripAdvisor page is more diverse than that of Sokos. The wordcount illustrates that 72 of the 111 English reviews extracted from Sokos hotels page were made by people from Finland. Comparatively, the number of reviews posted by Finnish people in English language on the page of Scandic was only 28. For Sokos the second largest geographical segment was USA with 8 reviews, followed by Singapore, which was mentioned only 4 times. The total number of different nationalities providing reviews in the given period for Scandic was 22 whereas for Sokos the number was 16. This finding seems to indicate that the customer base of Scandic is more international. Furthermore, the results seem to suggest that most of the guests at Sokos are locals and in-country travelers rather than tourists from abroad. On the other hand, this might suggest that international guests are more likely to post a review of their stay than domestic visitors. However, it should be noted that the findings only reflect the active reviewers on the online platform and therefore cannot be generalized to the whole customer base.

## 5. DISCUSSION AND ANALYSIS

To address the research question and objectives presented in the beginning, the study provided several interesting findings. Despite the small scale of the study and its limitation to only the very basic text mining methods, several findings support the claim that relevant information can, in fact, be extracted using text mining analytics. As discussed earlier in this study, customer engagement and active participation on online rating sites can strengthen a firm's competitive position by creating customer loyalty. Furthermore, it can enhance the company's image by positive peer-to-peer experience sharing and therefore create more business and increase profitability. The case study illustrates the potential of basic analytical tools for text mining analysis for the purpose of processing and analysing the content shared by consumers online.

In terms of consumer engagement, the findings show some interesting similarities and differences between the two hotels. Firstly, the average numerical ratings indicated that reviewers were generally slightly more satisfied with the service and quality of Sokos hotel, as it was more dominant of the two in excellent and good ratings. In addition, in the six-month period between September 2016 and February 2017 Sokos received more reviews than Scandic. However, the reviewers of Scandic were more active in expressing their positive feelings about all topics that were identified to be the most popular. As stated before in this study, it is proven that verbal customer reviews have more influential power over other consumers purchasing choices than the attached numerical ratings. This seems to suggest that although Sokos is performing stronger in terms of review frequency and ratings, it might have a weaker position in terms of the content of the verbal reviews. This finding was based on the frequency of occurrences of the most popular themes within the data. However, the sentiment analysis revealed slightly different results in terms of the opinions within the review content, which will be further discussed later in this section.

Secondly, the importance of responsiveness central in interaction between companies and consumers in online review platforms. According to Kwok and Xie (2016), the helpfulness of online hotel reviews is positively impacted by the responsiveness of managers and hotel staff to feedback and reviews. Furthermore, according to Xie, So and Wang (2017) a hotel's financial performance can be enhanced by providing timely and lengthy responses to online

feedback. Thus, a simple rate of response was calculated for both hotels. Comparing the number of reviews with official responses to those without responses, slight differences were encountered. Sokos had a slightly weaker score in responses as the findings seemed to suggest that the representatives of Sokos did respond to all reviews but in a less timely manner when compared to Scandic. Therefore, a minor improvement with the potential to strengthen the competitive position of Sokos would be to strengthen the frequency and speed of responses to the online reviews, as the literature seems to suggest that it might have a positive impact on financial performance.

The topics identified from the data were quite similar for both hotels. These could be categorized as location, service quality, breakfast and restaurant, general satisfaction, and rooms. In addition, a specific topic associated mostly with Scandic was the need for renovation. The results of topic classification illustrate the power of identifying underlying topics in the large unstructured mass of data. Although individual responses are required for reviews to keep customer satisfaction high, text mining applications can assist in mapping the overall trends by which decisions can be made. For instance, this case study illustrated word patterns that occurred together indicating satisfaction with various features. By doing so, managers are able to find the relevant discussion points present in the online content, to find out what consumers like and dislike about their services. The findings in this study illustrate that the reviewers of Scandic were more verbal in all categories, which could be seen from the number of occurrences of each identified topic in the data. In addition, the findings implied that the most common topics were similar for both hotels. Furthermore, the analysis discovered that many of the reviews for Scandic suggested the need for renovating the facilities. This kind of information is helpful for managers in discovering what consumers find as areas for improvement. Therefore, it assists in making strategic decisions in a responsive manner based on consumer demands and preferences.

Furthermore, the content was mined for sentiment. Opinion mining, also referred to as sentiment analysis, becomes a useful feature in text mining especially when consumer reviews are the subject of analysis. As stated by Alam, Ryu and Lee (2016), aspect based sentiment analysis is typically considered quite complex because the sentiment polarity varies according to the domain. This refers to the use of the same word in different context, where it refers to positive sentiment in the one and negative in the other. Therefore, the results of aspect based sentiment can be misleading in some cases. However, it can guide

managers as to the overall trends in opinions expressed online, which is why it seems useful to be demonstrated in this study. According to Liu (2012), opinions have been a subject of interest for companies and consumers for decades, because they have the ability to shape the thoughts and feelings of other consumers. Due to the emergence of online review platforms, both consumers and businesses can access the opinions expressed by previous guests, which can widely impact their decisions. Thus, even basic sentiment analysis tools become useful in conducting a competitive analysis on online ratings.

In the study, an aspect based sentiment analysis was conducted on with default settings of 15 aspects, which most often occur in hotel reviews. When looking at what competitive insights were derived from the sentiment analysis, the findings suggest that Sokos performed better in terms of the quantity of aspects with positive sentiment. From the data extracted on Sokos, positive sentiment was found in the following categories: View, Comfort, Food/Drinks, Room amenities, Cleanliness, Wi-fi, Beds, Location, Value, Staff, Design, and Quietness. As mentioned earlier, the results of sentiment analysis indicated a slightly stronger position for Sokos even though the topic classification analysis showed Scandic to have more occurrences of the popular topics. Scandic on the other hand had more aspects assigned with neutral sentiment, including Room amenities, Cleanliness, Beds, Value, and Design. Both data sets contained generally negative sentiment towards pricing. However, this can imply negative sentiment towards the general price level in area in which the hotels are located, and therefore the results cannot necessarily be generalized to depict the price range of the specific hotels. In addition, the room prices vary based on demand due to which the comments on pricing are dependent on the time of travel. However, the findings seem to suggest that Sokos has a more favourable position in terms of consumer satisfaction, as it performed more strongly in most categories with positive sentiment assigned.

The findings also illustrated the differences in the diversity of user backgrounds. A simple n-Gram data base illustrated how the majority of reviewers on the TripAdvisor page of Sokos were from Finland, whereas reviews on Scandic's page were from a more diverse variety of countries. Therefore, the differences in cultural dimensions might, in fact, influence the results derived from the reviews. Furthermore, as most reviews were posted in the second language of users, there might be language barriers that effect the content which was mined. As stated by Nguyen, Shirai and Velcin (2015), social media content often contains



misspellings and errors in grammar, which can affect the results of text mining analysis. Especially with the language constraints presented above, this should be noted in the analysis.

As consumer feedback is no longer private, it has power to shape the opinions and behaviour of other consumers more than ever before (He, Zha & Li, 2013). In addition, consumers have unprecedented power to influence the brand image of firms through spreading either positive or negative word of mouth online. Thus, the interaction on websites such as TripAdvisor bring aspects of the hotels' customer service into public inspection. This increasing accessibility of feedback brings new challenges in terms of mitigating negative opinions and providing responses that satisfy the wide audience that the online reviews reach. Thus, hotels should aim resources at tracking online conversation not only on their official channels but also popular third party platforms as demonstrated thorough this analysis.

Furthermore, companies can use the findings of text mining competitors' social media streams to implement competitive benchmarking (He, Zha & Li, 2013). By identifying what consumers are finding positive about a competitor's service, hotels can improve that aspect to create points-of-parity (POP). POPs refers to areas in which a company is performing strongly, due to which the competitor needs to handle it as strongly to be able to stay competitive (Krajnovic, Bosna & Jasic, 2012). For instance, if a hotel discovers consumers are dissatisfied with the beds in the hotel rooms but are admiring those of competitors, it might target resources to fix that problem to attract more consumers. If hotel companies are not aware of the strengths of competitors, they are more likely to lose customers to those companies. Furthermore, by recognizing the common trends present in the discussion about different players in the industry, hotels can find ways to differentiate themselves from competitors. Thus, hotels may create points-of-difference (POD), to create a competitive advantage and make their brand stand out from competitors. This is where automated content analysis comes into place, as it can assist in mapping the major points in social media discussions quickly and with little efforts from companies. Subsequently, companies can create POPs and PODs to strategically position themselves in the competitive market.

## 6. CONCLUSIONS

### 6.1. Main Findings

The main objective of this research was to find out the extent to which text mining can be beneficial in analysing hotel reviews, with the specific focus on Sokos and Scandic hotels in Finland. To explore this, reviews from TripAdvisor.com were extracted and analyzed to find the current trends and patterns present in the online discussions. By applying different text mining procedures, some relevant information was derived, which has the potential to be utilized to improve marketing operations and thus create more value to customers.

The main findings can be categorized to results from the quantitative data and the text mining findings from the qualitative data. The text mining results can further be categorized into topic classification and sentiment analysis. First off, the quantitative data present on the site revealed that Sokos had more reviews in the given period with a higher average. However, the rate of response to feedback given by guests was slightly better for Scandic, although both hotels performed quite well in this sector.

Furthermore, the text mining analysis uncovered the main discussion points of the reviews within the six-month period that was analysed. These topics were classified as location, service quality, breakfast and restaurant, general satisfaction, and rooms. For both hotels, the associations for these attributes were generally positive. In addition, it should be noted that in terms of frequency of occurrence of the major topics, Scandic had a leading position in all categories. These findings can help managers map the key trends in online discussion in order to emphasize the relevant points in marketing in addition to making other strategic decision based on consumer feedback. The analysis also found that there were several reviews implying the need for renovation especially in the case of Scandic. This also demonstrates the power of finding whether constructive feedback appears in large quantities making it a trend more than an individual suggestion. This enables firms to react to the feedback in a more effective way.

In terms of sentiment, the study adduced opinions relating to 15 of the most common aspects of hotel reviews indicating whether they were positive, neutral, or negative. The

aspect based sentiment analysis illustrated that Sokos received positive sentiment in more categories than Scandic, whereas Scadic remained neutral in several categories. As sentiment is one of the most powerful shapers of consumer opinions and purchasing decisions, opinion mining can benefit companies greatly to determine the major trends in consumer sentiment. Overall, the findings illustrated that basic text mining procedures can, indeed, be used to derive relevant trends and patterns from large quantities of unstructured data for the basis of making strategic decisions.

## 6.2. Implications for International Business

As the significance of travel review applications is growing and the availability of text mining technology is increasing, new opportunities in the field open constantly. Although the case study was conducted in hotels located in Finland, the basics of text mining online reviews apply to hotels regardless of location. In addition, the hotel industry is often international to begin with, as the rise of tourism brings customers from different countries world-wide. Thus, there are several implications of this study for international business.

Firstly, the research seems to suggest that by screening hotel reviews and finding the most relevant discussion points within the data, hotels can find its strengths and weaknesses from the consumers' perspective. Furthermore, this is not limited to only finding out the trends in the feedback for a hotel itself but also identifying them in the content available about competitors. Automated text analysis technologies do, in fact, enable environmental screening in a more efficient and effective way, which can help hotels position themselves in the international market place. In addition, companies can use the findings of text mining analysis on competitors' social media streams for their advantage, by utilizing competitive benchmarking (He, Zha & Li, 2013). This means that firms can map what competitors are doing strongly with the aim of copying those aspects into their own operations. Furthermore, although the hotels on which the competitive analysis was conducted in this case are located in Finland, the fundamental idea behind to study and the objectives of this research can be applied to any hotel regardless of location. The RapidMiner software, for instance, has features to process text in English, German, Czech, Arabic and French. In addition, it enables the use of self-made dictionaries which allows users to create

dictionaries in their own language. Thus, the software can be used to process content in several different languages, opening opportunities internationally for business use.

Secondly, there are already several options for free text mining software online, one of which was used for this case study. This enables companies to use text mining as a basic tool for content analysis, without the need for great financial investments or high-end sophisticated technologies. In addition, the study demonstrates how basic text mining techniques can be used to conduct a content analysis on social media data without the need of extensive technological knowledge or expertise. Thus, companies should familiarize themselves with such options world-wide, to strengthen their position in the international business environment. As demonstrated, even basic text mining procedures can offer several insights to large quantities of consumer generated content, which can benefit companies in various ways. Therefore, text mining has the potential to become a common tool for content analysis in the hotel industry internationally.

### 6.3. Suggestions for Further Research

There are several aspects related to digital marketing and text mining in the hospitality industry that could be further studied in the future. As the scope of this study was quite small, the results cannot be generalized. Therefore, a more comprehensive study could be performed in the future with a larger sample of reviews. In addition, the scope of the research could be broadened to process the hotel replies as well as the reviews, in order to map how efficient hotels are at responding to feedback. In addition, the text mining procedures used in the study were very basic and simple, which gives the opportunity of applying more complex features to gain more insights on the topic.

One of the findings was the distribution of nationalities among the users posting reviews on the TripAdvisor website. As Sokos is a Finnish hotel chain and Scandic is foreign in the Finnish context, it would be interesting to see if that impacts the amount of English reviews posted about them. In addition, it seems that Sokos is more successful at targeting local guests whereas the customer base for Scandic seems more diverse according to the sample in the case study. Therefore, to gain more in-depth insights on the hotels studied in

this case, research could be conducted on the marketing efforts that the hotels do to reach their target customers. In addition, the research was limited to include only reviews posted in the English language. Therefore, future research could be conducted to reviews in other languages to identify how linguistic and cultural differences may influence the content of hotel reviews. In addition, as the current study focused merely on the content of the travel rating websites, it did not take into account financial aspects of the subject hotels performance. An area for future research could add this perspective to give more comprehensive results of the possible effects on financial performance.

## REFERENCES

Abubakar, A. & Ilkan, M. (2016) 'Impact of online WOM on destination trust and intention to travel: A medical tourism perspective.' *Journal of Destination Marketing & Management* [Online]. 5 (3): 192-201.

Ady, M. & Quadri-Felitti, D. (2015) *Consumer research identifies how to present travel review content for more bookings*. Available from: <http://www.trustyou.com/press/study-reveals-travelers-prefer-summarized-review-content-full-text-reviews> [Accessed on 6 January 2017]

Agag, G. & El-Masry, A. (2016) 'Understanding consumer intention to participate in online travel community and effects on consumer intention to purchase travel online and WOM: An integration of innovation diffusion theory and TAM with trust.' *Computers in Human Behavior* [Online]. 60: 97-111.

Alam, M. H., Ryu, W. & Lee, S. (2016) 'Joint multi-grain topic sentiment: modeling semantic aspects for online reviews.' *Information Sciences* [Online]. 339: 206-223.

Amaro, S. & Duarte, P. (2015) 'An integrative model of consumers' intentions to purchase travel online.' *Tourism Management* [Online]. 46: 64-79.

Arroyabe, A. M. (2016) 'Factors affecting social interaction on social network sites: the Facebook case.' *Journal of Enterprise Information Management* [Online]. 29 (5): 630 – 649.

Bjorkelund, E., Burnett, T. H. & Norvåg, K. (2012) 'A study of opinion mining and visualization of hotel reviews' In: *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services*; Bali, Indonesia: 3-5 December. pp. 229-238.

Cenni, I. & Goethals, P. (2017) 'Negative hotel reviews on TripAdvisor: A cross-linguistic analysis'. *Discourse, Context & Media* [Online]. 16: 22-30.

- Cesaroni, F. M & Consoli, D. (2015) 'The Adoption and use of Social Media by Micro and Small Enterprises.' In: *Proceedings of the 2nd European Conference on Social Media*; Porto, Portugal: 9-10 July.
- Chung, N. & Koo, C. (2015) 'The use of social media in travel information search.' *Telematics and Informatics* [Online]. 32 (2): 215–229.
- Clark, J. (2012) 'Text Mining and Scholarly Publishing.' *Publishing Research Consortium* [Online]. 1-20.
- Fuggle, L. (2017) *70 travel & tourism statistics to know about in 2016*. Available from: <https://www.treksoft.com/en/blog/travel-tourism-stats-2016> [Accessed on 15 March 2017].
- Gémar, G. & Jiménez-Quintero, J. A. (2015) 'Text mining social media for competitive analysis.' *Tourism & Management Studies* [Online]. 11 (1): 84-90.
- Guo, Y., Barnes, S. & Jia, Q. (2017) 'Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation.' *Tourism Management* [Online]. 59: 467-483.
- Harrysson, M., Metayer, E., & Sarrazin, H. (2012) 'How social intelligence can guide decisions.' *McKinsey Quarterly* [Online]. (4): 81–89.
- He, W., Zha, S. & Li, L. (2013) 'Social media competitive analysis and text mining: A case study in the pizza industry.' *International Journal of Information Management* [Online]. 33 (3): 464-472.
- Hicks, A. (2013) 'Digital Marketing in an Outreach context.' *University Libraries Faculty & Staff Contributions*. 25.
- Hu, N., Koh N. S. & Reddy, S. K. (2013) 'Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales.' *Decision support systems* [Online]. 57: 42-53.

Hu, Y., Chen, Y. & Chou, H. (2017) 'Opinion mining from online hotel reviews – A text summarization approach.' *Information processing & Management* [Online]. 53 (2): 436-449.

Jannach, D., Zenker, M., & Fuchs, M. (2014) 'Leveraging multi-criteria customer feedback for satisfaction analysis and improved recommendations.' *Information Technology & Tourism*. 14 (2): 119-149.

Khadjeh Nassirtoussi, A., Aghabozorgi, S., Ying Wah, T. & Ngo, D. (2014) 'Text mining for market prediction: A systematic review.' *Expert Systems with Applications* [Online]. 41 (16): 7653-7670.

Kim, K., Park, O., Yun, S. & Yun, H. (2017) 'What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management.' *Technological Forecasting and Social Change*. [Online]

Kwok, L. & Xie, K. (2016) 'Factors contributing to the helpfulness of online hotel reviews.' *International Journal of Contemporary Hospitality Management* [Online]. 28 (10): 2156-2177.

Lagrosen, S. O. & Grundén, K. (2014) 'Social media marketing in the wellness industry.' *The TQM Journal* [Online]. 26 (3): 253-260.

Lau, K., Lee, K. & Ho, Y. (2005) 'Text mining for the Hotel Industry.' *Cornell Hotel and Restaurant Administration Quarterly* [Online]. 46 (2): 344-362.

Leong, E., Ewing, M. & Pitt, L. (2004) 'Analysing competitors' online persuasive themes with text mining.' *Marketing Intelligence & Planning* [Online]. 22 (2): 187-200.

Liu, B. (2012) 'Sentiment Analysis and Opinion Mining.' *Synthesis Lectures on Human Language Technologies* [Online]. 5 (1): 1-167.

Liu, Y., Teichert, T., Rossi, M., Li, H. & Hu, F. (2017) 'Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews'. *Tourism Management* [Online]. 55: 554-563.



Mankad, S., Han, H., Goh J. & Gavirneni, S. (2016) 'Understanding Hotel Reviews Through Automated Text Analysis.' *Service Science* [Online]. 8 (2): 124-138.

McDonald, D. & Kelly, U. (2012) *Value and benefits of text mining*. Available from: <https://www.jisc.ac.uk/reports/value-and-benefits-of-text-mining> [Accessed on 4 January 2017].

Nguyen, T. H., Shirai, K. & Velcin, J. (2015) 'Sentiment analysis on social media for stock movement prediction.' *Expert Systems with Applications* [Online]. 42 (24): 9603-9611.

Pauwels, K., Aksehirli, Z. & Lackman, A. (2016) 'Like the ad or the brand? Marketing stimulates different electronic word-of-mouth content to drive online and offline performance.' *International Journal of Research in Marketing* [Online]. 33 (3): 639-655.

Qi, M., Li, X., Zhu, E. & Shi, Y. (2017) 'Evaluation of perceived indoor environmental quality of five-star hotels in China: An application of online review analysis.' *Building and Environment* [Online]. 111: 1-9.

Reimer, T. & Bankenstein, M. (2016) 'When good WOM hurts and bad WOM gains: The effect of untrustworthy online reviews.' *Journal of Business Research* [Online]. 69 (12): 5993-6001.

Scandichotels.com. (2017) *THE LARGEST HOTEL NETWORK IN THE NORDICS*. Available from: <https://www.scandichotels.com/weekend-packages-and-offers/largest-hotel-chain-nordics> [Accessed on 16 March 2017].

Sokoshotels.fi. (2017) *Sokos Hotels lyhyesti*. Available from: <https://www.sokoshotels.fi/fi/tietoa-meista> [Accessed on 16 March 2017].

Sun, A., Lachanski, M. & Fabozzi, F. (2016) 'Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction.' *International Review of Financial Analysis* [Online]. 48: 272-281.

Sun, J., Wang, G., Cheng, X. & Fu, Y. (2015) 'Mining affective text to improve social media item recommendation.' *Information Processing & Management* [Online]. 51 (4): 444-457.

Upshall, M. (2014) 'Text mining'. *Business Information Review* [Online]. 31 (2): 91-99.

Liau, B. Y. & Tan, P. P. (2014) 'Gaining customer knowledge in low cost airlines through text mining.' *Industrial Management & Data Systems* [Online]. 114 (9): 1344-1359.

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* [Online]. 58: 51-65.

Xie, K., So, K. & Wang, W. (2017) 'Joint effects of management responses and online reviews on hotel financial performance: A data-analytics approach.' *International Journal of Hospitality Management* [Online]. 62: 101-110.

Xun, X. & Li, Y. (2016) 'The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach.' *International Journal of Hospitality Management* [Online]. 55: 57-69.