Swansea University
Prifysgol Abertawe

Management Responses to Online Reviews: Big Data From Social Media Platforms

Submitted to Swansea University in fulfilment of the requirements for the Degree of

Doctor of Philosophy

By

Aytac Gokce

School of Management

Swansea University

2022

Abstract

User-generated content from virtual communities helps businesses develop and sustain competitive advantages, which leads to asking how firms can strategically manage that content. This research, which consists of two studies, discusses management response strategies for hotel firms to gain a competitive advantage and improve customer relationship management by leveraging big data, social media analytics, and deep learning techniques. Since negative reviews' harmful effects are greater than positive comments' contribution, firms must strategise their responses to intervene in and minimise those damages. Although current literature includes a sheer amount of research that presents effective response strategies to negative reviews, they mostly overlook an extensive classification of response strategies. The first study consists of two phases and focuses on comprehensive response strategies to only negative reviews. The first phase is explorative and presents a correlation analysis between response strategies and overall ratings of hotels. It also reveals the differences in those strategies based on hotel class, average customer rating, and region. The second phase investigates effective response strategies for increasing the subsequent ratings of returning customers using logistic regression analysis. It presents that responses involving statements of admittance of mistake(s), specific action, and direct contact requests help increase following ratings of previously dissatisfied returning customers. In addition, personalising the response for better customer relationship management is particularly difficult due to the significant variability of textual reviews with various topics. The second study examines the impact of personalised management responses to positive and negative reviews on rating growth, integrating a novel method of multi-topic matching approach with a panel data analysis. It demonstrates that (a) personalised responses improve future ratings of hotels; (b) the effect of personalised responses is stronger for luxury hotels in increasing future ratings. Lastly, practical insights are provided.

Keywords: management responses, online reviews, negative reviews, social media, big data analytics, natural language processing, deep learning

Dedication

To my mother, Sunay Bozkurt, and my family for their eternal love, support, and encouragement during my journey to complete this thesis.

Acknowledgement

I am thankful to Allah (God) for giving me the courage, patience, and strength to complete this PhD thesis.

I would like to express my deepest gratitude to my supervisors, Dr. Nick Hajli and Dr. Thomas Roderick, whose sincerity, motivation, and encouragement I will never forget. I am thankful for their guidance, valuable comments, and constructive suggestions during my PhD. I also thank my previous supervisor Dr. Mina Tajvidi for her time to support me.

I am grateful for my beloved mother whose constant love and support keep me motivated and confident. The biggest reason for this success is to make her a little proud in return for the effort she put into raising three children by herself and bringing them to this day. My deepest thanks to my siblings Cagri and Korkut, who keep me grounded, remind me of what is important in life, stood by me in my most difficult times, and are always supportive of my adventures. Finally, I owe my deepest gratitude to Ezgi, who is my love. I am forever thankful for the unconditional love and support throughout the entire thesis process and every day.

Finally, my thanks and appreciations go to the Republic of Turkey Ministry of National Education for funding my PhD studies.

Declaration and Statements

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed [redacted]

Date: 10/06/2022

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s). Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed [redacted]

Date: 10/06/2022

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed [redacted]

Date: 10/06/2022

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans after expiry of a bar on access approved by the Swansea University.
Signed [redacted]

Date: 10/06/2022

Table of Contents

List of Figures

List of Tables

Abbreviations

AEs	Autoencoders
CNN	Convolutional Neural Network
CRM	Customer Relationship Management
DBN	Deep Belief Network
eWOM	Electronic Word of Mouth
IE	Information Extraction
IT	Information Technology
KBV	Knowledge-based View
NER	Named Entity Recognition
NLP	Natural Language Processing
PoS	Part of Speech
RBM	Restricted Boltzmann Machine
RNN	Recurrent Neural Network
SMA	Social Media Analytics
SVC	Support Vector Classifier
WOM	Word of Mouth

CHAPTER 1:

INTRODUCTION

1.1 Research Background

The amount of data available is constantly increasing due to technological advancements in the Internet of Things era. The unprecedented production of data has led the term “big data” to become a phenomenon in recent years. The concept of big data evokes the exponential growth and complexity of raw data that is broadly characterized by volume, variety, velocity, veracity, and value (Fosso Wamba et al., 2015; Gandomi & Haider, 2015; White, 2012). As a result, recognition its significance has highly increased across industries and sectors (Gandomi & Haider, 2015). A reason why big data is attractive is the presence of hidden patterns where so much valuable information can be found (Choi et al., 2017). Thus, it has been seen as a critical resource that can help firms to generate substantial business insights (Elhoseny et al., 2020) and improve decision-making (Wamba et al., 2017). In addition, to survive and thrive within an immensely competitive and rapidly changing environment, developing sufficient sources and capabilities to use big data analytics efficiently is critical to sustaining one’s competitive advantage (Erevelles et al., 2016; Shan et al., 2019).

In the last decade, social media has become an important source of big data with the overwhelming generation of unstructured data in a short timescale from diverse platforms and websites that contain several types of content such as posts, reviews, tweets, and comments (Lyu & Kim, 2016; Sebei et al., 2018). Managing the information from the huge flow of social media data for business goals is a new issue in business-focused knowledge management (He et al., 2017). As such, generated insights from big data analyses of social media allow marketers to better understand online communities, predict consumer behaviours, and make related business strategies (Ghani et al., 2019; He et al., 2017). However, the creation of big data with social media makes the social media analytics process a more demanding task in which traditional techniques lack efficiency in analysing this large social media data (Hayat et al., 2019). This gives rise to the need for advanced technologies to improve performance and achieve reliable results for the implemented analysis. In particular, deep learning, a subset of machine learning, paves the way for social media big data analytics (Hayat et al., 2019). Due to its considerable efficiency in processing massive and unstructured data such as images, text, and speech (Shin, 2020).

Using technologies like natural language processing and computer vision with deep learning algorithms could help organisations develop meaningful market outcomes and uncover new strategies in real time.

In line with social media, big data concepts, and the rapidly changing business environments with growing interest in big data analytics, this research aims to suggest that organizations can obtain valuable business insights with data-driven strategies using novel techniques. Being positioned in big data from social media, the research investigates how businesses can use the combination of deep learning techniques and big data analytics to boost their competitive capacity, strategic decision-making, and managerial efficiency in the business environment**.** On this basis, this research is conducted to help management in their decision-making and reaction to social media posts by indicating strategic content generation**.**

1.2 Research Scope and Motivation

The extensive use of social media platforms in recent years has led to the rapid generation and exchange of user generated content (UGC), which emerges by users as a result of their background details and daily activities such as connecting others, viewing, and engaging in other users and their activities (Ayeh et al., 2013). As an important source of big data, UGC can be in several forms: online reviews, posts, images, video/audio files, and blogs that are teeming with consumer behavioural insights and promise unprecedented potential opportunities to businesses for sustainable sustainability and dynamic competitiveness. For instance, electronic word of mouth (eWOM), a form of UGC, has become a powerful marketing tool (Reimer & Benkenstein, 2016; Weisfeld-Spolter et al., 2014). Consumers are used to participating in social networking sites with the motivation of finding information or sharing their ideas (Hollebeek et al., 2014; Levy & Gvili, 2015). eWOM is described as “dynamic and ongoing information exchange process between potential, actual, or former consumers regarding a product, service, brand, or company, which is available to a multitude of people and institutions via the Internet” (Ismailova et al., 2017, p. 18). Numerous studies report the influence of eWOM on customers’ perceptions and behaviours linked with making a purchase decision (Chen et al., 2015; Elwalda et al., 2016;

Tsao & Hsieh, 2015) and on business performance (Kim et al., 2019; Nieto et al., 2014; Xie et al., 2016).

Considering the increment in the dynamic and continuous information flow of the online crowd and its potential effects, effective management of communications in social networking platforms is now highly important for the development and implementation of proactive business strategies. Accordingly, firm engagement, also known as business engagement, has gained a broad interest from researchers and industry in recent years (Bai & Yan, 2020; Sheng, 2019; Yang et al., 2016). It signifies that businesses are not only cognizant of the importance of online social interactions but also acting as a part in the social networking sites by being involved in communications. Engagement in social media can underpin businesses in several ways, such as promoting product sales, enhancing brand image and loyalty, improving brand awareness, and reducing marketing costs (Bai & Yan, 2020; Kim et al., 2015; Wan & Ren, 2017). Additionally, firms can monitor and analyse consumers' attitudes and sentiments to discover knowledge and absorb new ideas, and they can generate and deliver messages to consumers. As a result, active management and participation in online interactions can lead to the generation of strategic insights that can help decision-making (Sheng, 2019).

A popular type of firm engagement in online social interactions is management responses to customer reviews, which enables firms to proactively engage in social networks rather than being passive listeners (Xie et al., 2017). Management response can relay valuable information about the quality of services with unobservable attributes that prospective customers can consider and evaluate to adjust their expectations before making purchase (Sheng, 2019). For firms, it is a useful tool to manage customer relationships by interacting with customers who decide to post positive or negative reviews at the post-consumption stage (Gu & Ye, 2014). Considering the fact that negative reviews are likely to reach a wider audience and damage business more than positive ones promote (Chevalier & Mayzlin, 2006; Cui et al., 2012; Hornik et al., 2015), responding to negative reviews is recognised as a critical intervention approach in a case of service failure (Piehler et al., 2019). Thus, managers are recommended to provide appropriate and strategic responses to protect a firm’s reputation and alleviate potential adverse effects. These strategies are undoubtedly about more quantitative such as response ratio, speed, and length. The quantitative features are considered signals of management responses, which lack the heterogeneous nature of the response content and ignore the spontaneous associations between responses and reviews (Li et al., 2020). The content might be the most crucial component of management responses, delivering information to customers regarding how their concern is handled and what the reasons for the problem are (Li et al., 2020). Research explores the response content with a simple classification of management responses, accommodativeness vs defensiveness, (Lee & Cranage, 2014; Lee & Song, 2010; Li et al., 2018), response sentiment, (Sheng, 2018), showing empathy, (Min et al., 2015), and explanation vs meta discourse (Li et al., 2020) without a detailed exploration of response content. Although other existing studies concentrate on the effects of various response strategies, they widely overlook a comprehensive analysis of the content of responses to negative reviews with the help of big data analytics and advanced techniques. Therefore, the first phase of Study–1 in this research aims to fill this gap with an explorative analysis by focusing on an in-depth investigation of managerial response strategies to negative online reviews.

In addition, most of the literature investigates the effectiveness of managerial responses to on online reviews that are posted by all customers. Nevertheless, managers respond to the customers who visit the hotel only once may not have a major impact on subsequent ratings (Sheng et al., 2019). Wang and Chaudry (2018) mention that the inadequacy of returning customers in their sample restricted them from examining user heterogeneity. Research proves the positive influence of response provision on subsequent ratings of repeated customers (Gu & Ye, 2014; Sheng et al., 2019). However, there is a gap in the literature about what strategies could be more influential in increasing subsequent ratings of returning dissatisfied customers. In light of the first phase of Study–1 and focusing on response content, this second phase of Study–1 aims to fill the gap in the literature by concentrating on effective response strategies to increase subsequent ratings of repeated dissatisfied customers

Furthermore, several studies show the importance of customising management response content based on the review (Li et al., 2018; Wang & Chaudhry, 2018; Zhang et al., 2020).

However, there is limited evidence in the literature demonstrating the effect of personalised management responses in future ratings. Zhang et al. (2020) prove that personalised responses can increase subsequent ratings of customers by calculating the topic matching degree between responses and reviews using machine learning techniques. Inspired by their research but different from their multiclass text classification approach that allows responses and reviews to only have one matched topic, Study–2 in this thesis aims to complement their study by deploying multilabel text classifications with deep learning technology and calculating multitopic matching degrees between responses and reviews. Compared to Study–1 in this thesis and several other studies that focus on responses to only negative reviews, Study–2 covers management response strategies applicable to both negative and positive reviews and focuses on whether personalised responses influence subsequent ratings.

The motivation for carrying out this research is that the subject of management responses is still open to development and the belief that knowledge-based view and dynamic marketing capabilities can significantly assist firms in the hotel industry in developing appropriate engagement strategies in online social networks. In this vein, this research seeks how generated knowledge through social media big data can help managers strategise their reactions to reviewers and improve customer relationship management, hence gaining a competitive advantage.

1.3 Research Questions and Objectives

1.3.1 Research Aim

The aim of this research is to investigate and reveal the strategic content of management responses to customer reviews in service failure and to explore whether personalised management responses are influential on subsequent ratings of customers. To do so, the research is conducted with two studies by collecting publicly available data from TripAdvisor, which is considered one of the largest travel platforms with more than 450 million monthly unique users and over 600 million reviews of hotels, restaurants, and other relevant businesses (TripAdvisor, 2018). Study–1 includes the explorative examination of responses to negative reviews and the examination of effective response strategies that might increase subsequent ratings of previously dissatisfied repeated customers. Study–2 involves examination of the effect of the topic matching degree between responses and online reviews (considering both positive and negative) on rating growth.

1.3.2 Research Questions

To fulfil the research aim, there are three key questions that guide this research:

Q1: What strategies do managers follow when responding negative online reviews?

Q2: Which response strategies are more practical to increase the subsequent ratings of returning dissatisfied customers?

Q3: Do personalised responses with a multitopic matching degree affect subsequent ratings?

1.3.3 Objectives

The objectives of this research are presented as follows:

- To review literature based on management responses to online reviews and analyse quantitative and qualitative strategies.
- To explore qualitative response strategies to negative online reviews by adopting advanced data analytics.
- To identify effective response strategies on the ratings of returning dissatisfied customers.
- To investigate the effect of personalised responses on rating growth, revealing topic matching degree between responses and reviews (including both positive and negative reviews).

1.4 Research Methodology

The research consists of two major studies to accomplish the aims and objectives. Overall, this thesis adopts a mixed-methods approach. Mixed-methods research benefits from both quantitative and qualitative methods by gathering, analysing, and merging qualitative and quantitative data in a single project (Bryman & Bell, 2015). The collected data for this thesis comprises unstructured textual and numeric data. The qualitative textual data are quantified with an intense textual annotation and classification process. Therefore, the adopted philosophy in this research is considered as pragmatism.

Both studies are conducted using social media and big data analytics, natural language processing, and machine and deep learning techniques. The first phase of the first study explores the management response strategies to online negative reviews with the support of data visualization techniques. To reveal management's response strategies to negative reviews in the first phase of the first study, a set of randomly selected responses are labelled with a binary annotation approach based on five predefined categories in the data annotation stage. Then various machine and deep learning algorithms are tested, and the models with the best accuracy are selected as binary classifiers for each response category separately. These models are applied to the large data set separately, and a data set is created for each response category. Later, these datasets are combined, and a single dataset showing all the categories included in responses is created. The data set is then analysed using descriptive statistics and data visualization techniques. The second phase of the first study examines the effective response strategies to previously dissatisfied returning customers. Because this phase focuses only on returning dissatisfied customers, the data are relatively smaller. Hence, all the responses are manually labelled without the need of a text classification algorithm. Logistic regression analysis is then applied to reveal the strategic responses that can be effective in increasing the next rating of returning customers.

Unlike the first study, the second study involves managerial responses to both negative and positive reviews to investigate the effect of personalized management responses on subsequent ratings of customers. The data annotation and classification technique in this study are also different from the first study because customer reviews (positive and negative) are annotated in addition to management responses for the purposes of capturing the number of matched topics between a response and corresponding review. The second study adopts a multilabel text annotation and classification approach as opposed to the first study where binary text annotation and classification are adopted. With the multilabel classification approach, all responses and reviews are allowed to be assigned to more than one topic at the same time. Finally, multilevel regression is adopted in the data analysis stage, and the effect of topic-matching degree on subsequent ratings is analysed monthly.

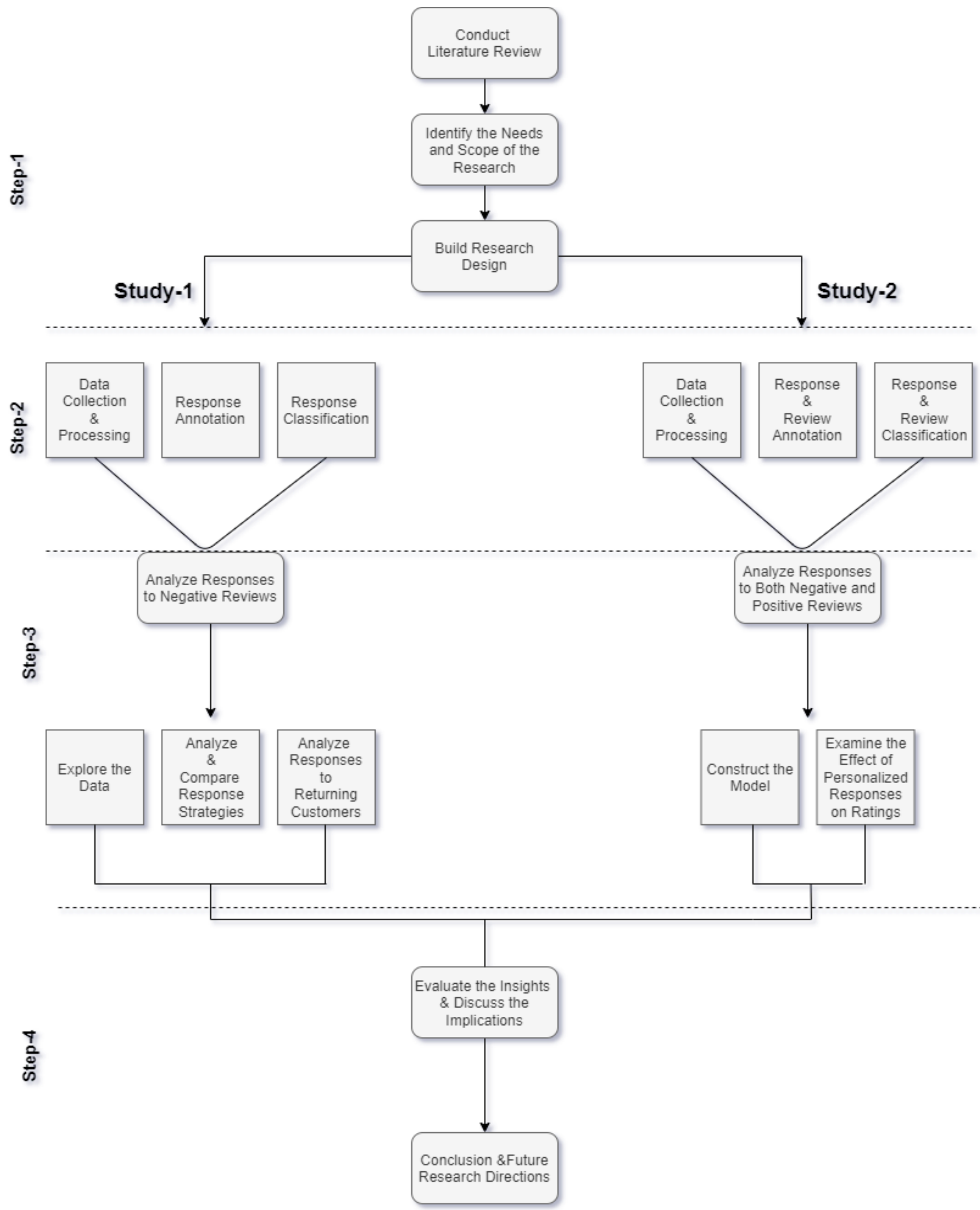
1.5 Research Design

Figure 1 shows the research design implemented in four main steps for the two studies. In the first step, a comprehensive literature review of management responses and social media big data analytics is conducted. Subsequently, the scope and needs of the research are determined, and the research is designed for two independent studies. In the second step of Study–1, management responses to negative reviews are collected, processed, annotated, and classified. The third step of Study–1 consists of an exploratory analysis with descriptive statistics and numerous data visualizations that are carried out to reveal what strategies managers are following when responding to dissatisfied customers (the first phase of Study-1), plus a logistic regression analysis to examine what strategies are more effective in increasing subsequent ratings of returning dissatisfied customers (the second phase of Study-1).

In the second step of Study–2, both positive and negative reviews and corresponding management responses are collected, processed, annotated, and classified. The third step of Study–2 involves the calculation of topic matching degrees between and responses and reviews, construction of the model, and investigation of whether higher topic matching degree leads greater rating growth on subsequent ratings of customers in a monthly timeframe.

In the last step of the research, the results are evaluated and discussed. After all, the research is finalized with the conclusion and future directions.

Figure 1 Research Design



1.6 Research Contributions

This research initially contributes to big data literature by social media research in the scope of business and management. Instead of using traditional approaches such as surveys or interviews, this research adopts big data analytics from social media for a large amount of qualitative and quantitative data. The mixed-method design and advanced techniques transform huge information from online social interactions into observable instruments, thereby allowing a more extensive analysis of the research problems (Shen, 2018). Using publicly available big data on social media to capture particular aspects of customer opinions and business reactions, this research provides actionable insights into the strategic firm engagement with online social interactions within the e-business context.

In particular, the first study contributes to the management response literature on negative reviews. As a form of intervention tool, management responses play a significant role in dealing with negative reviews (Istanbulluoglu, 2017; Ma et al., 2015). Using a large amount of social media data and advanced analytics, Study–1 in this research explores how managers strategizes response content for negative reviews, with the provision of a comprehensive comparative analysis of response strategies based on hotel class (star) and overall ratings. There are also few studies in the literature about managerial responses to dissatisfied returning customers. The second phase of this study fills this gap by revealing effective response strategies to increase subsequent ratings of dissatisfied returning customers.

Moreover, the second study in this research complements the literature on personalized management responses. Studies prove that tailoring managerial response to corresponding reviews is significant in many aspects. For instance, Wang and Chaudry (2019) state that personalized management responses to the negative reviews have positive effects on subsequent customers. Studies also show that specific responses addressing issues or showing empathy improve the trust of potential customers and the quality of communication (Min et al., 2015; Wei et al., 2013). However, there is limited research about the efficacy of personalized management responses to both positive and negative reviews on subsequent ratings. More recently, Zhang et al. (2020) demonstrate that personalized responses have a positive effect on hotel rating increases. Based on the five common topics, which they define with a topic modelling approach, taking part in customer review content, they calculate the probability that management response involves the same topic as the corresponding review. However, their approach allows each review and response to involve only one topic. In fact, many customer reviews frequently contain more than one topic where they mention favourable and unfavourable aspects of the service they experience. The second study in this research fills this gap by investigating multitopic matching degrees between responses and reviews. In doing so, the approach in this study allows a review and response to contain more than one topic if it exists in customer reviews and management responses. The results also reveal that higher multitopic matching degrees have a positive effect on hotel rating increases.

The findings from this thesis support the concept of the dynamic marketing capabilities model. In marketing, IT plays a pivotal role in dynamic capabilities to exploit competitive advantage through service excellence and customer intimacy (Mikalef & Pateli, 2017). Obtaining marketing information through data mining and analysis of consumer generated information has been shown to support businesses to better understand what is desired in the market and take appropriate action to improve service, thereby, improving performance and increasing customer satisfaction. Additionally, this research highlights effective communication styles when interacting with customers to improve customer relationship management. It also supports the concept of dynamic marketing capability, emphasizing the combination of social media and digital marketing strategies to enhance relationship management capabilities between businesses and customers (Wang & Kim, 2017). In addition, this research expands current research on firms' knowledge base by investigating how firms in the highly dynamic and competitive hotel industry leverage large data from social media to generate new knowledge and build new strategies in responding to customers through big data and advance analytics for favourable business outcomes.

This research has major practical contributions. This research suggests managers to engage in online social interactions and follow appropriate strategies. Responding to dissatisfied customers with the statements including specific action is taken for the related issue, accepting the responsibility for problems occurred, and direct-contact request could increase subsequent rating of dissatisfied customers. Additionally, this research suggests to managers that providing of personalized responses to online reviews by tapping into many of the topics presented in customer reviews can increase customers' subsequent ratings.

1.7 Outline of the Thesis

This chapter has presented a brief background of the research through followings: research scope and motivation, aims and objectives, research methodology, and research contributions. The remainder of the thesis is outlined as follows.

Chapter 2 provides a review of the related literature and identifies research gaps. More specifically, this chapter first presents an overview of eWOM and its comparison with WOM. The chapter then carries on with a survey of social media and its relation with customer relationship management. It then continues with a comprehensive review of online management responses to customer reviews. The remaining part of this chapter involves more technical part related to social media big data analytics. In this stage, social media data analytics techniques and methods, natural language processing, and advanced techniques including a brief of machine learning and deep learning are introduced. Finally, the chapter introduces the related theories of this research.

Chapter 3 outlines the adopted research methodology and the design of the research. Differences of research methodologies and research approaches are presented in this chapter, and then adopted research methodology and research approach are introduced with relevant reasons.

Chapter 4 provides Study–1. A background of management responses to online negative reviews is initially discussed, and the need of this scope is identified here.

Chapter 5 presents Study–2. A background of personalized management responses is discussed, and the relevant research gap is identified.

Chapter 6 presents the conclusion of this research. It first presents a summary of findings from the two studies. This is followed by a discussion on the theoretical contributions as well as practical implications on management responses to the customers' reviews scope. The last section of this chapter involves future directions and research limitations.

CHAPTER 2:

LITERATURE REVIEW

2.1 Introduction

This chapter initially presents a brief review of WOM, then the emergence of eWOM and differences between these two concepts are introduced and an overview of social media is presented. The transformation of traditional customer relationship management into social customer relationship management is then handled with the benefits it provides to organizations when used effectively. After introducing these concepts, the specific scope of this research, management responses, is reviewed in the literature. Management responses to negative reviews and personalized management responses are not introduced in this chapter. Instead, as two different studies, they are separately presented in Chapter 4 and Chapter 5, respectively.

The second part of this chapter involves technical concepts regarding social media big data analytics. It initially presents an overview of social media big data analytics. Thereafter, it briefly introduces social media analytics techniques consisting of social network analysis, sentiment analysis, text mining, and natural language processing and discusses advanced analytics with a survey of machine learning and deep learning techniques.

The chapter ends with indicating supported theories which are dynamic capabilities and knowledge-based theory.

2.2 Theories

2.2.1 Dynamic Capabilities

The dynamic capabilities theory is mainly an expanded version of the resource-based view of firms (Teece et al., 1997). The resource-based view signifies that the firms' sustainable competitive advantages are determined by their nonhomogeneous resources (Barney, 1991; Barney et al., 2011). Notwithstanding, the theory is challenged in the current dynamic environments and urges scholars to extend the resource-based view to dynamic capabilities (Gutierrez-Gutierrez et al., 2018; Teece et al., 1997). The term “dynamic capabilities” was originally put forward by Teece et al. (1997), for whom the term the term referred to an organizations ability to integrate, build, and reconfigure internal and external competencies to address rapidly changing environments. Dynamic capabilities are associated with the capacity of the adequate and timely adaption to the altering environment with resources and reconfiguration of external or internal processes together with the existent competencies (Eisenhardt & Martin, 2000; Gaur et al., 2014). Consistent with this perspective, effective reconfiguration and transformation necessitates a capability to comprehend relevant environmental changes, continuous observation of the market and technology, and enthusiasm for the adoption of pertinent practices and benchmarking (Braganza et al., 2017). As such, the more often organizations reconfigure and transform their capabilities, the more likely it is to accomplish a competitive advantage (Braganza et al., 2017). Moreover, the notion of dynamic capabilities underlines not only the growth of management capabilities, but also combinations of technological, organizational, and functional capacities, which are difficult to imitate (Teece et al., 1997). Unlike the resource-based view, which suggests the usage of managerial practices to generate new abilities, the dynamic capabilities view proposes the usage of management strategies to rebuild competencies depending on the alterations in the environment (Shamim et al., 2019).

Dynamic capabilities involve features related to effective processes (Eisenhardt & Martin, 2000) that generate collaborative networks between various internal and external relationships to create source combinations to live up to or exceed stakeholders' expectations (Hill et al., 2014). Some common examples of dynamic capabilities include product development processes, knowledge creation processes, and the resource allocation processes. Organizations join technologies, data, abilities, and expertness to design products and services or create greater efficiencies. Considering all these, dynamic capabilities stimulate organizations to continuously generate new thinking and to innovate (Braganza et al., 2017).

Dynamic capability emphasizes firms' strategic planning in leveraging opportunities and their capacity to restructure business competencies in changing market conditions, customers, and technologies (Teece, 2007). Particularly, Barrales-Molina et al. (2014) describe the dynamic marketing capabilities as a combination of absorptive capacity and knowledge management in the acquisition of market knowledge to improve organization. Utilizing knowledge from the market can help companies to develop dynamic marketing capabilities regarding adapting to market conditions, developing product and services, detecting customer behaviour and needs, and enhancing strategies (Barrales-Molina et al., 2014; Morgan, 2011). Studies in dynamic capabilities show that firms with powerful dynamic capabilities have an advantage in a competitive environment due to their unique skills to detect and react to new market information (Junni et al., 2015; Teece et al., 2016; Weber & Tarba, 2014). Within this regard, it is proposed by some that the current big data era calls for firms to be more adaptable and proactive in recognizing threats and facilities, as well as make use of any possible opportunities using resources like consumer reviews and online blogs to sustain their innovation and to enhance competitiveness (George et al., 2014; Teece et al., 2016). With the aid of online review data analysis, businesses can perceive customers' behaviour and new opportunities which enable them to both redesign product and service offers and reconfigure their business models (Sheng et al., 2019). Leveraging user generated data tends to allow businesses to innovate and respond to customers' demands in the changing business environment (Wessel, 2016).

2.2.2 Knowledge-Based View

According to the knowledge-based view (KBV), knowledge is the most valuable resource for firms aiming to build new competitive advantage sources (Nonaka, 1994; Robert, 1996). "Knowledge" here stands for any skill, belief, or information that is applicable by organizations to their activities (Vikas et al., 2002). A knowledge-based view of the firm is based on the notion that a firm's achievement is related to the extent to which it is able to advance its knowledge-base by either generating or attaining new knowledge, integrate its discrete knowledge domains, and use its knowledge in improvement or development of products or processes (Kogut & Zander, 1992; Nonaka, 1994; Robert, 1996). A firm's KBV also proposes that the capability of the firm to be in a better position than its rivals is estimated on its capacity to enhance or gain advantage from valuable knowledge by learning (Sheng et al., 2019). Learning tends to provide firms with captured insights, which can help them to boost performance (Bogner & Bansal, 2007).

By identifying and comprehending market opportunities, firms can develop solid market orientation cultures to be prosperous in competitive environments (Atuahene-Gima, 1995).

In this regard, gaining knowledge by deploying social media big data (Sheng et al., 2019) tends to allow firms to develop specific knowledge that is hard to simulate or exploit for innovation (Argote & Ingram, 2000). On the strength of this perspective, some suggest that firms possessing upper abilities in seizing big data for knowledge development and transfer might have more potential to have superiority over their competitors (George et al., 2014; Wamba et al., 2017). Regardless of which industrial domain their scope is in, the willingness of the firms has been increasing in deploying big data based sources to create a competitive advantage thanks to the feasible efficacies derived from big data analytics (George et al., 2014; Thomas, 2013). With the aid of insights generated from big data analytics, firms can figure out emerging situations and reposition themselves properly (Wamba et al., 2017). In addition, insight generated through big data analytics can enable firms to develop the sheer volume of knowledge and help firms in expanding the locus of decision-making (Mikalef et al., 2020), product innovation, and customer responses (Akter et al., 2016).

Several knowledge management initiatives concentrate on obtaining, examining, and exploiting customer information (Braganza et al., 2017). Considering the growth of big data analytics applications in the organizational context, it has been indicated to achieve the comprehension of new unprecedented business opportunities by capturing customer data from the combination of different data sources (Braganza et al., 2017). Grasping knowledge with text mining and NLP techniques of online customer reviews are highly crucial for the extremely competitive hospitality industry. Firms in this industry can enhance their service quality, as well as customer satisfactions, which tends to increase performance and overall ratings (Liu et al., 2018; Sheng et al., 2019; Xie et al., 2016). In addition to that, capturing knowledge from organizations' responses, which this research aims at, can put some recommendations with respect to identifying effective response strategies for organizations.

2.3 Word-of-Mouth

The word-of-mouth (WOM) concept is a particular form of social impact on customer behaviours (Lis & Neßler, 2014) regarded as an informal communication amongst noncommercial consumers regarding products and services (Anderson, 1998; Arndt, 1967; Harrison-Walker, 2001). The power of WOM is based on its proneness to rapidly evolve and spread without the contribution of the needs of the firms.

The existent studies have widely brought out the significance of WOM with respect to consumers' decision-making (Anderson, 1998; Trusov et al., 2009; Zhu & Zhang, 2006). Prior to making purchasing decisions, consumers seek knowledge from other buyers to be able to get enlightened or support what they already know about goods and services (Berger, 1988; Lim & Chung, 2011). According to Trusov et al. (2009), the effect of WOM is far more than other marketing actions as well as media events.

The impact of WOM is greater than conventional communication tools on customers' attitudes and the decision-making process (Lis & Neßler, 2014; Sen & Lerman, 2007). As such, the dissemination of the information via consumers is seen as more reliable and realistic considering that it does not have a marketing purpose (Sen & Lerman, 2007). From the customer perspective, the goods and services' quality standard tend to be more assessable considering other consumers' selections as guides in their decision-making. Particularly, WOM is a convenient information source for consumers to learn something negative regarding what they are willing to purchase (Lis & Neßler, 2014). Similarly, Keaveney (1995) points out that a vast majority of customers inform each other while they switch their product or service provider and clarify the motive of the change. Therefore, the influence of WOM is indisputably remarkable in the marketplace.

2.4 Electronic Word-of-Mouth

The emergence of Web 2.0 evolved WOM to electronic word-of-mouth (eWOM) (Mauri & Minazzi, 2013). The power of the internet in information dissemination in an expeditious manner added a new dimension to WOM. The comprehensive definition of eWOM in the marketing literature refers to continuing an active exchange of information amongst prospective, actual, or previous customers with respect to services, products, brand, or company which is reachable by a lot of people and organizations via the internet (Hennig-Thurau et al., 2004; Ismagilova et al., 2017). On the basis of the definition, the most evident distinction between WOM and eWOM is the type of information transfer.

Due to the multitude of recipients on the internet, eWOM is remarkably more influential than traditional WOM in many aspects. As opposed to traditional WOM, eWOM communication may occur in various tools such as emails, blogs, review websites, and social networks sites (SNS) (Cheung & Thadani, 2010). Social media, with its platforms and websites, is indeed highly convenient for eWOM, so that consumers now have more control over their media behaviours and play more active roles in the process of decision-making regarding products and services of brands. Social networking sites especially have turned into popular online communication channels with their visually and textually enriched contents. Social networking sites enable consumers to generate visible profiles, create networks, and have interpersonal comments openly by wiping time and geographical limitations out. Thus, customers are easily and rapidly able to interact with each other and have access to global audiences to be enlightened about the related products or services. With the help of online reviews and ratings, consumers now have the opportunity to leave and look for sentiments or ratings regarding the goods and service comments of people they do not know at all, thanks to anonymous and dynamic features of the internet (Goldsmith & Horowitz, 2006).

Consequently, there are a number of attributes of eWOM put forth by researchers; the range and volume of eWOM are unsurpassed (Dellarocas, 2003), so that scores of consumers can communicate in a short span of time, thereby enabling eWOM to raise a larger awareness by comparison with traditional WOM (Kiecker & Cowles, 2002). Another important characteristic of eWOM is its permanency and observability in different platforms for other customers seeking opinions regarding what they want to purchase (Dellarocas & Narayan, 2007). This feature indicates that present eWOM affects future eWOM. By considering these characteristics, eWOM's influence on customers' attitudes, purchase decisions, as well as brand choices is unavoidable.

eWOM may be positive or negative (Lis & Neßler, 2014), as with classic WOM, which may result in both advantages and challenges for organizations. Customers have the opportunity to publicly articulate their satisfaction about the related matter, which can positively influence potential customers who seek comments before making the purchase. As such, positive eWOM tends to promote a company, product, and services as an efficient tool to attain new customers and retain current ones (Dellarocas, 2003; Luo & Homburg, 2007). Research also indicates the relation between positive eWOM and sales performance.

However, social media allows customers to engage in negative eWOM to voice their negative experiences with respect to brands, products, and services when things do not live up to their expectations. Marketing research points out that negative eWOM has the strongest destructive impact on a provider among post failure activities (Azemi et al., 2020; Choi & Choi, 2014; Wang et al., 2011). Numerous studies mention that negative eWOM has a huge potential to encourage other customers to get involved in the complaints, showing the wide reach of complaints (Gu & Ye, 2014; Obeidat et al., 2017; Rosenmayer et al., 2018). Moreover, negative eWOM likely leads to a decrease in trust, loyalty, and repurchasing the customers' inclination (Jalilvand & Heidari, 2017; Umashankar et al., 2017). Negative eWOM might also be detrimental to companies’ reputation and earnings (Dellarocas, 2003; Hennig-Thurau & Walsh, 2003; Lis & Neßler, 2014). Nevertheless, negative eWOM can help companies better understand customers’ reactions to and expectations of their products or services. Because of these reasons, companies should take heed of eWOM and know how to gain advantages from it to conduct appropriate proactive strategies (Ismagilova et al., 2017).

2.4.1 Online Reviews

Online customer reviews are a form of e-WOM communications with no structural constraints and are presented in free-form prose(Bai & Yan, 2020). They typically contain a wealth of information, such as purchasing experience, vendor details, customer complaints, user experience, satisfaction scores, and consumer ratings for various goods and services (Jin et al., 2019). They have described consumer experiences in greater detail than standard rating surveys and can therefore more accurately reflect customer sentiments (Zablocki et al., 2018). For many applications, such as sentiment analysis(Chen et al., 2015), product ranking prediction (Li et al., 2011), and summarising customer evaluations (Amplayo et al., 2022), scientific research has been done to assess textual reviews. Other analyses attempted to assist commercial applications, such as service quality evaluation (Palese & Usai, 2018) and hotel aspect rating prediction(Wang et al., 2010).

The main components of online consumer reviews are numerical ratings and descriptive comments (Fang et al., 2016). The characteristics of online reviews are frequently regarded as valence, volume, and variance. Review valence is the positive or negative orientation of information about an object or situation (Buttle, 1998). Typically, positive comments can lead to positive attitudes and strong buying intent, whilst negative comments can lead to negative attitudes and low intent to purchase (Zablocki et al., 2018). East et al.(2008) discover that positive reviews increase purchase likelihood. Ye et al.(2009) also find that the amount of favourable reviews correlates highly with an increase in sales. Furthermore, positive reviews enhance customers' perceptions of a hotel at the pre-consumption stage (Cheung & Thadani, 2012; Ladhari & Michaud, 2015), which has an advantageous effect on their purchasing decision (Ajzen et al., 2018). In addition, positive communications boost customers' trust in the hotel (Sparks & Browning, 2011), which ultimately increases their intention to book or make a purchase (Plotkina & Munzel, 2016). A number of research have demonstrated that negative evaluations affect purchasing decisions more than positive ones (Chang & Wu, 2014; Cui et al., 2012; Yoo et al., 2013). According to Hennig-Thurau et al. (2004), bad evaluations are a powerful instrument that significantly influences consumers' perceptions of a firm and its brands. Negative information tends to be more diagnostic, valuable, and enlightening than positive information; thus, it is given greater weight in decision-making processes (Bambauer-Sachse & Mangold, 2011).

Review volume denotes the number of online reviews or ratings for a particular product or brand (Chintagunta et al., 2010; Floyd et al., 2014). Higher review quantity correlates with increased product awareness and, thus, increased sales (Nikolay et al., 2011). Consumers are more persuaded by products or companies with numerous online evaluations, as the perceived accuracy of an opinion grows when a large number of consumers shares it (Salganik & Watts, 2008). Additionally, many internet reviews can enlighten buyers about a product, which in turn drives product sales (Salganik & Watts, 2008).

Variance reflects reviewers' disagreement regarding a product or service, as seen by a spectrum of positive and negative words (Minnema et al., 2016). Variance harms customers' buying decisions (Babic Rosario et al., 2016; Floyd et al., 2014) since it diminishes expectations and heightens uncertainty due to a broader range of favourable and unfavourable judgments (Chen & Lurie, 2013; Khare et al., 2011). Zhu & Zhang (2010) demonstrate that a large coefficient of variation reduces product sales. According to Wang et al.(2015), consumers are also more likely to disregard a product if it is presented in a context with a high level of variation, as they hesitate that it may not meet their requirements and preferences.

2.5 Social Media

Indeed, the internet is the backbone of societies in the current era. With the technological innovations bringing the internet to the various types of devices such as computers, smartphones, wearables like smartwatches, and many others, users are now able to share or create content worldwide (Tuten, 2018). Currently, there are almost 4.8 billion internet users that account for much more than 50% of the total population, and most of them take part in at least one social network (Kemp, 2019). The numbers demonstrate that not only does information diffuse by corporations and governments but also flows across the people. The ultimate change in the way information spreads is based on the evolution of the web. In Web 1.0, the interactions between users were quite limited so that the users were mostly able to reach static contents of the internet. Likewise, Web 1.0 can be seen as a portal of information where users passively attain information without having the facility for active interactions like posting comments, reviews, and feedbacks. In addition to networks of information, Web 2.0 added dimension to the internet by connecting people's networks (Murugesan, 2007). Tim O'Reilly (2005) who popularized the term, expresses Web 2.0 as services and sites that based upon the generation of content by users instead of editors and content editors (O'Reilly, 2005). As this implies, Web 2.0 enables interactive data flow between users. It also facilitates access to rich data and encourages participation and information sharing (Tuten, 2018).

The creation of Web 2.0 gave rise to the emergence of the concept of "social media." Despite the interchangeable usage of the terms "Web 2.0" and "social media," the meaning of the term social media is different from the meaning of Web 2.0 (Constantinides, 2014). The definition of social media can be viewed from a few different perspectives. Practically, it is an accumulation of software-based digital technologies that usually exist as applications and websites that allow users to connect with each other by sending and receiving information or content over social networking sites. Within this context, we can think of social media as platforms and their characteristics. However, Appel et al. (2020) defines that social media can be seen less as a particular technology services and digital media but more as digital places in which users canalize important part of their lives. In this respect, social media is less relating to platforms or technologies and more relating to what people do in these digital environments. More broadly, social media is a concept that comprehends anything (such as people, information, content, behaviours, and organizations) available in interconnected and networked digital environments in which interactivity exists (Appel et al., 2019).

Social media has now turned into one of the most important components of digital technologies of our lives with several platforms and billions of users throughout the world. Sharma and Verma (2018) categorised social media according to their primary purpose. They characterised social media as a set of platforms, including social networks, media sharing, discussion forums, bookmarking, and blogging. Social networking sites (SNS), a subdomain of social media, are a networked communication platform where individuals 1) have uniquely identifiable profiles consisting of user-supplied content, content provided by other users, and system-provided data; 2) can publicly articulate connections that can be viewed and traversed by others; and 3) can consume, produce, or engage with streams of the user-generated content offered by their connections (Ellison & Boyd, 2013). SNS often involve forming and maintaining personal and professional online interactions via various venues (Alhabash & Ma, 2017). Facebook is one of the most popular social networking sites, serving approximately 3 billion users worldwide (DataReportal, 2021). It enables individuals to interact with their friends, family, and acquaintances and post and share information such as images, stories and status updates(Alhabash & Ma, 2017). With around 450 million users (DataReportal, 2021), Twitter, as a SNS, is a microblogging site where users engage in real-time by sending 140-character tweets to their followers. Users can engage in conversation with mentions, responses, and hashtags (Alhabash & Ma, 2017). Instagram is a photo and video sharing platform that enables users to capture photos, record videos, apply filters, and publish them (Alhabash & Ma, 2017). The number of Instagram users has also exceeded one billion in 2021 (DataReportal, 2021). Online discussion forums provide users with various topics for questions and comments. The underlying principles by which a group of people develop collective thought may be better understood by considering how people behave in online forums (Medvedev et al., 2020). Reddit, for instance, is among the most popular online platforms, with around 430 million users (DataReportal, 2021). Social bookmarking networks such as Pinterest and Flipboard allow users to add, annotate, and manage web pages while surfing the Internet for information from diverse sources (Duong, 2020). Additionally, users can share these sites with others in order to collectively investigate topics of interest. A blog is a personal website published about a particular subject. It allows users to share their enthusiasm with the world and fosters an active community of readers who comment on the author's articles (Duong, 2020). Tumblr is currently one of the largest blogging sites in the world, with over 500 million blogs and almost 14 million daily entries (Tumblr, 2020). Strikingly, estimating the aggregate number of social media users will be above 45% of the world's population in 2024 (eMarketer, 2020). Considering the extraordinary number of users who spend a nonnegligible amount of their time in social media along with a wide variety of platforms, social media has unsurprisingly become a highly effective marketing channel by marketers.

Social media, which has tended to be considered as a form of eWOM in marketing (Appel et al., 2019), has been widely recognized as an influential notion which makes contributions to the businesses' marketing objectives and strategies, particularly in customers' participation, customer relationship management, and communication (Filo et al., 2015; Saxena & Khanna, 2013). One of the vital features of social media in marketing is to serve as a bridge between customers and companies, thereby, enhancing two-way communication as well as involving customers more with the brands (Alalwan et al., 2017). This becomes possible through social media's capability to present the information or various content types such as visual, verbal, and textual or a mixture of some of or all these types (Okazaki & Taylor, 2013). From the customers' perspective, social media-based communication enables them to find what they are seeking in an easy and rapid manner, reducing their efforts to search for information (Laroche et al., 2013). Plus, they can express their satisfactions or dissatisfactions with the brand, product, or services. On the other side, firms have important opportunities by taking advantage of social media in many of their interactions with consumers. The uses include improving customer service, managing customer relationships, promoting, and enhancing their product or services, developing new products, and managing brand reputations (Alalwan et al., 2017; Tuten, 2018).

2.6 Social Media and Customer Relationship Management

With the enhancement in technology and the internet, customer relationship management (CRM) has gained another perspective, as there used to be one-way communication in which firms transmitted their offline marketing methods to online platforms by not making use of the potential of the internet with traditional technologies. Today, CRM has been growing more than ever with the prevalence of social media. Social media has turned into an appropriate platform for CRM and its fundamental relationship to marketing principles with its interactive and relational features (Choudhury & Harrigan, 2014; Hennig-Thurau et al., 2010).

Generally, CRM is regarded as an integration of technological resources involving tactical and operational aspects within the organization and other strategic aspects such as customer engagement and customer orientation to attain important performance gains (Borges et al., 2009; Chang et al., 2010; Coltman, 2007). The notion of CRM concentrates on systems' supporting the elements like sales, marketing, analysis, and data integration. The emergence of social media technologies transformed the CRM concept by adding such innovations to performance. Greenberg (2010) describes social CRM as "a philosophy and a business strategy, supported by a system and a technology, designed to engage the customer in a collaborative interaction that provides mutually beneficial value in a trusted and transparent business environment"(Greenberg, 2010). The definition shows social CRM is similar to traditional CRM in that it involves the accepted concept of traditional CRM by adding social entities that comprise the interactions between customers and customers and businesses to customers. It also indicates in the definition that new technologies and systems consist of supporting interaction with customers with the aim of maintaining long-term relationships with improved performance (Choudhury & Harrigan, 2014). As such, social CRM is based on two-way communications and interactive relationships with the customer in which they play an important role in the cocreation of marketing efforts (Rodriguez et al., 2012).

Being able to create and maintain social ties between their customers and brands usually allow organizations to have a close and strong relationship with those customers (Alalwan et al., 2017). In this sense, organizations broadly adopt social media to contribute to CRM and customer experience. Using social media actively and creatively, organizations can positively affect the interactivity and association with customers. Researchers also suggest that effective social media usage in CRM brings out the potential to be influential for organizations' performances, as the nature of social media facilitates more customer engagement, interactions, and information sharing (Brodie et al., 2013; Hennig-Thurau et al., 2010; Raj et al., 2012).

Social CRM has also drawn huge attention by organizations to overcome crisis situations. As abovementioned, not all customer interactions are positive. When something goes wrong, most social customers do not hesitate to share their poor experiences on open social platforms. The frustrations customers experience can be spread out rapidly and viewed by a high number of users and hurt an organization's image. That is why it is one of the top priorities of organizations is to take account of customer complaints seriously in social CRM (Tuten, 2018). Organizations must have strategies to make up for setbacks and regain dissatisfied customers. As such, organizations are encouraged to listen to conversations of customers, observe those conversations, and act on those conversations to manage mishaps that customer experience (Tuten, 2018).

2.7 Management Responses

With the great number of review websites, such as TripAdvisor, Yelp, and Foursquare, customers are now able to write and share their experiences with respect to products or services. For instance, if a customer has a satisfied or dissatisfied experience staying at a hotel or dining at a restaurant, they prefer to post a review online on social media platforms rather than complaining to the firms' customer service (Sheng et al., 2019). Customers' tributes to or damnation of products or services shared on social platforms are public and may be viewed by millions of others, and these reviews are the most significant information source when customers make a purchase decision (Litvin et al., 2008). Therefore, online reviews' potential accessibility and mass communication urge businesses to seek how to benefit from user-generated context to gain insights and take appropriate action in the competitive environment (Marshall et al., 2015; Sorescu, 2017).

Formerly, businesses maintained a competitive advantage responding to customers' positive or negative comments directed at them (Sheng et al., 2019). Even though the response helped stand and develop in the past, because of the existence of many social media sites, firms cannot solely depend on their websites to provide a higher quality of service and satisfy customers' expectations (Constantiou & Kallinikos, 2015; Sorescu, 2017). In today's data-rich environment, online customer reviews have remarkable potential to affect businesses. While positive ones can attract prospective customers and enhance businesses' performance, negative ones can impair businesses' online reputation and performance (Lappas et al., 2016; Wang & Chaudhry, 2018). As such, managers are suggested to devote close attention to online customer reviews with the goal of developing effective strategies to enhance business performance and brand reputation by proactively influencing eWOM opinions of customers (Gu & Ye, 2014; Xie et al., 2014).

Management responses, "which take the form of an open-ended piece of text and are publicly displayed underneath the consumer review being addressed" (Xie et al., 2017), have been paid rising attention as the review of relevant literature indicates. Generally, several researchers conduct experiments to reveal the probable positive influence of managerial responses on customers' satisfaction (Gu & Ye, 2014; Istanbulluoglu, 2017; Min et al., 2015; Zhang et al., 2020), trust (Sparks et al., 2016; Wei et al., 2013), engagement (Li et al., 2017), business reputation (Lee & Cranage, 2014; Lee & Song, 2010; Levy et al., 2013; Min et al., 2015; Sparks & Bradley, 2017), and hotel sales (Kim et al., 2015; Li et al., 2018; Mauri & Minazzi, 2013; Xie et al., 2017; Xie et al., 2014).

By looking at the attributes of management responses in the literature, volume, speed, match rate, and content of responses are common attributes that play important roles in management responses. It is worth it to point out that the amount of management responses displayed in the review part is the most ostensible feature of management responses on the website, which can lead the reviewers and readers to have a perception of how a hotel is responsive to customers' dissatisfactions (Sheng et al., 2019). Empirical studies demonstrates that there exists a positive relationship between the volume of management responses and customer ratings (Proserpio & Zervas, 2017; Sheng, 2018; Xie et al., 2016; Zhang et al., 2020), also business performance (Xie et al., 2017).

The service recovery literature highlights that a service failure is more probably to be successfully solved if the problem is promptly tackled. As such, a timely response is not only essential to customer evaluations of service quality but also effective in increasing perceptions of fairness. Increasingly, studies investigate the role of response speed in the online context. Sparks et al. (2016) find that a prompt response to a negative review enhances customer inferences of trustworthiness. Sheng (2018) also demonstrates that there is a positive association between the speed of response and subsequent customer ratings. Li et al. (2017) show that speedy responses significantly affect customers' engagement, leading to an increase in reviews and popularity ranking. In contrast, Min et al. (2015) report that the speed of the response to negative reviews does not influence customer ratings.

Response length, referred to the number of words in the response, can reveal the amount of information conveyed in the response, so that it can play a crucial role in affecting customers' decisions. Such conclusion arises from uncertainty reduction theory which defines that the amount of information is influential in communication outcomes by decreasing ambiguity. In this case, the provision of a management response with lengthy descriptions can help prospective customers perceive to a wider extent the hotel's approach on how to take care of customer concerns. Additionally, a detailed management response provides customers to clearly evaluate the incident according to an additional source to what other customers said. Research examines the effect of response length on various parameters. Li et al. (2017) find that the response length has a nonsignificant effect on customer engagement. Shen (2018) states that response length has no significant impact on future ratings. Xie et al. (2017) demonstrate there is a positive association between response length and hotel performance for 4-star and 5-star hotels.

In terms of match rate between management responses and online reviews, Sparks et al. (2016) point out that when managers provide relevant responses and solutions about customers' reviews and concerns, customers might receive the affirmative signal that the management team attach importance to their ideas. Following, customers might adopt a positive attitude towards a hotel and therefore prefer to stay in the hotel, increasing hotel performance. More recently, Zhang et al. (2020) apply text mining techniques to investigate management responses from a topic matching perspective, and they demonstrate that management responses with high level topic matching spearhead a rise in overall hotel rating. Furthermore, Sheng et al. (2019) examine management responses to online reviews by looking from another perspective that they concentrate on the impact of responses in subsequent ratings of repeated customers by using a text analytics approach. Their study highlights that repeated customers' subsequent ratings are higher if responses are provided to their previous reviews.

Several studies explore the relative effectiveness of the response content. Many studies concentrate on accommodative (taking responsibility, apologizing, and expressing regret) and defensive responses (denying the responsibility, rationalizing). Accommodative responses perform more effectively (Lee and Son) and effective in enhancing prospective customers' attitudes when given to high consensus reviews and increasing sales when given to product failure reviews. However, defensive responses are found effective when given to low consensus reviews, when given in an analytical format, and when the negative review is ordinary.

Moreover, studies demonstrate the influence of adapting the response content based on the corresponding reviews. Wang and Chaudry (2019) define that personalized management responses to the negative reviews have positive effects on subsequent customers. Studies have also shown that specific responses addressing issues or showing empathy improve the trust of potential customers and quality of communication (Min et al., 2015; Wei et al., 2013). More recently, Zhang et al. (2020) explore the impact of personalized management responses to both positive and negative reviews on customer satisfactions using a topic matching method and found that responses with a higher level of topic matching give rise to an increase in subsequent ratings.

Other features of responses are examined in the literature. Li et al. (2020) introduce two features (meta discourse and explanation) of response content to negative reviews, and they demonstrate that explanations are more effective than meta discourse for reviews involving anger whereas meta discourse outperforms explanations for reviews involving anxiety. Researchers also investigate particular response strategies including apology, explanation, action, acknowledgment, correction, and several other form of strategies (Leung et al., 2013, Levy et al., 2013, Sparks and Bradley, 2017).

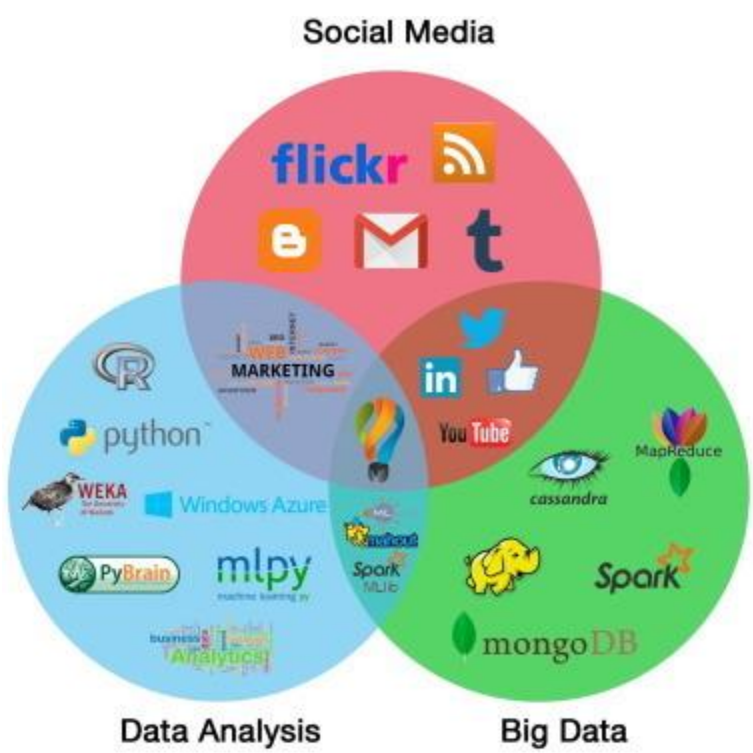
2.8 Big Data and Social Media

The contents of the web, from publisher to user-created content, have evolved due to the Internet and Web 2.0 technologies' enhancements (Alexander, 2006). It is now easily possible to access information such as topics, trends, sentiments, and reviews in a global manner from the Internet. Furthermore, the dissemination and use of social media have given rise to immense opportunities and challenges for practitioners and researchers (Holsapple et al., 2018). Social media platforms make the exploitation of massive data valuable for a variety of fields, including politics (Stieglitz & Dang-Xuan, 2012), the biomedical field (Kotsilieris et al., 2017), business decision-making process (Rahmani et al., 2014), and customer relationship management (Ballings & Van den Poel, 2015). Consequently, the advent of social media analytics as a process (Gandomi & Haider, 2015) tries to extract meaningful information from social data by collecting, cleaning, and analysing the data. In its complete form, social media analytics (SMA) refers to the typical multi-process work of analysing social media-related data (Sebei et al., 2018). Examples of SMA techniques include sentiment analysis (Li & Wu, 2010), sentiment classification (Xu et al., 2020), social network analysis (Yang et al., 2018), and data mining (Patil & Kulkarni, 2018), which rely on a variety of associated techniques, including text mining, computational linguistics, machine learning, and natural language processing.

According to the Digital 2020 October Global Statshot Report, more than 4 billion people worldwide now use social media each month, as well an average 2 million new users joining them. Social media usage by billions causes the generation of overwhelming data, and social media has turned to one of the most relevant and representative data sources for big data. Big data creation via social media has occurred with the platforms such as social networking sites, blogs, microblogs, media sharing, wikis, question and answer sites, and reviews sites, which involve the important pieces of information concerning user interactions, daily activities, and sharing (Holsapple et al., 2018). This information has spread to many different areas such as everyday life such as e-commerce, e-business, e-tourism, hobbies, and friendship (Kaplan & Haenlein, 2010) education (Tess, 2013), health (Farsi, 2021), and daily work. The exponential growth of heterogeneous data (videos, photographs, texts and audio) makes the social media analytical process difficult, mainly due to the inefficiency of traditional techniques to analyse this enormous flow of social media data (Bello-Orgaz et al., 2016). That increases the need for new technologies to assist in boosting and improving the performance of conventional procedures and generating accurate results based on the implemented analysis (Peng et al., 2017). These innovative solutions are essentially Big Data technologies that, when combined with conventional social media analytics, demonstrate outstanding potential for processing social media data (Bello-Orgaz et al., 2016). Defined by Bohlouli et al. (2015), He et al. (2017), and Bella-Orgaz et al. (2016), *Social Big Data (SBD)* or *Big Social Data (BSD)* has emerged to describe this new field of study as a combination of two domains: social media and big data. SBD refers to combining Big Data technologies and frameworks with traditional analysis methodologies to process and analyse social media data to derive value (Sebei et al., 2018). Similarly, Nguyen et al. (2015) and Vatrapu (2016) define SBD as the social media data generated and characterised by a vast amount of unstructured data. This term is also used as *Social Media Big Data* in the literature (Bukovina, 2016; Ghani et al., 2019).

The conceptual depiction of the three fundamental social big data fields is depicted in Figure 2. Social media as a data source, big data as a parallel and large processing paradigm, and data analysis as a collection of algorithms and methods used to obtain and analyse knowledge(Bello-Orgaz et al., 2016). The intersections between these clusters represent the idea of combining these notions. For instance, the intersection of big data and data analysis presents some machine learning frameworks. Machine learning is a subfield of artificial intelligence that various social media sites have utilised to recognise data trends(Ghani et al., 2019). Classification(Aggarwal, 2015; Reuter & Cimiano, 2012), clustering (Lim et al., 2017), and deep learning are among the most prevalent techniques (Asgari-Chenaghlu et al., 2021; Pathak et al., 2019). The intersection of data analysis and social media exemplifies the concept of Web-based applications that extensively use social media data. The intersection between big data and social media is displayed in some social media platforms utilising big data technologies (MongoDB, Cassandra, Hadoop, etc.) to create their Web platforms. The centre of this diagram depicts the primary objective of any social big data application: the extraction and use of knowledge.

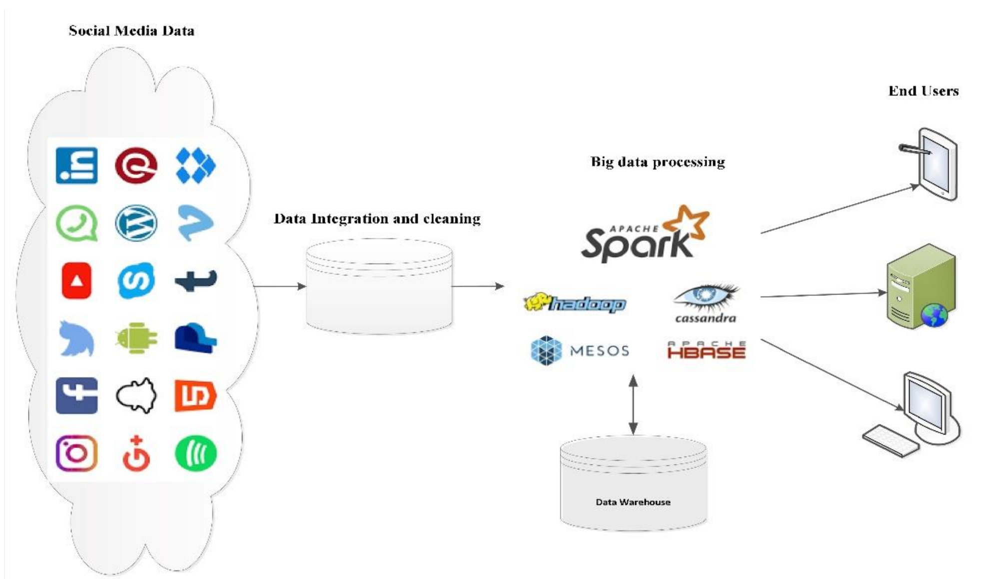
Figure 2 The conceptual map of Social Big Data



Source: Bello-Orgaz et al. (2016).

Notwithstanding, dealing with massive data acquired from social media in different formats has also presented some issues (Ghani et al., 2019). When not correctly used to inform decision-making by transforming a vast volume of social data into insightful knowledge, big data gathered from social media are meaningless (Gandomi & Haider, 2015). The social media big data pipeline, which is proposed by Ghani et al. (2019), is depicted in Figure 3. First, the data are gathered from multiple social media sources and stored in big data storage systems that can accommodate massive amounts of information, like HDFS, Hbase, and Cassandra. Before leveraging technologies like Spark, Hadoop, and Mesos to process large data, data integration and cleansing are used. Consequently, users can examine the outcome of data processing on numerous devices, including PCs, servers, and smartphones.

Figure 3 Social media big data processing (Ghani et al., 2019).



Source: Ghani et al. (2019)

2.8.1 Big Data

Data is continuously generated at an ever-growing rate in today's technology era. The rise of social media, mobile phones, the Internet of Things (IoT), multimedia, and healthcare technologies has led to an immense flow of data in different formats (Gerard et al., 2014). Likewise, the advancement in open-source frameworks, such as Hadoop, has been one of the key roles in developing big data, whereby implementing big data is easier to work with and store (White, 2015). The data creation at an extraordinary level is referred to big data, and its growth has emerged as a phenomenon and been paid attention to deploy by organizations, governments, and academia to create and capture values in many application areas for their growth (Manyika et al., 2011).

The term ‘big data’ is attributed to the mid of 1990s, according to Diebold et al.(2012), which was first used by retired former Chief Scientist Silicon Graphics John Mashey. However, after many years, the term has become a buzzword thanks to the promotional initiatives and considerable investments in building big data analytics platforms of leading companies, such as IBM and Oracle (Gandomi & Haider, 2015). Generally accepted, big data is a term used for vast and complex databases which are impossible to process and handle by traditional systems and data warehousing tools. Notwithstanding, there is no agreed definition of big data, like many terms used to refer to rapidly evolving technologies (Rob, 2014). Even though big data evoke size as the first characteristic, big data has several characteristics, with volume, variety, and velocity being well-known(Laney, 2001; Zikopoulos et al., 2013). For example, Gartner, who originally introduced these three terms, defines big data as ‘*Big data is high volume, high velocity, and/ or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization*’(Laney, 2001). In the same way, Mills et al. (2012) describes big data as ‘*Big data, a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.*’.

Both definitions denote the three fundamental characteristics of big data: Volume, Variety, and Velocity. By adding a fourth "V," Value, other organizations and big data practitioners (such as academics, engineers, and others) have expanded the 3V model to become a 4V model (Hashem et al., 2015). In addition, if the notion of veracity is added to the definition of big data, this model could be expanded to the five Vs. This set of characteristics provides a concise and widely agreed definition of what constitutes a big-data-based problem, application, software, or framework and what does not. The following is a brief description of these concepts (Laney, 2001):

Volume denotes the magnitude of created data surpassing terabytes and reaching petabytes and even exabytes. Data are continuously generated from various sources, such as social media, cloud-based services, enterprise-related data, and data linked to the IoT. Based on an estimation made by International Data Corporation estimation in 2018, data creation will jump to 175 zettabytes in 2025 from 33 zettabytes (Rydning, 2018).

Variety refers to different forms of data available (Chen & Zhang, 2014). Big data can consist of multiple types of data containing structured, semi-structured, and non-structured. *Structured data* refers to the highly organized data having a defined style, format and structure, usually found in spreadsheets and relational databases. *Semi-structured data* is a form of structured data that does not comply with the inherent structure of data models in relational databases or other data tables; however, it contains patterns such as tags or markers to be parsed. Extensible Markup Language (XML) and JavaScript Object Notation (JSON) are examples of semi-structured data forms. *Unstructured data* stands for the data with no predefined data model and is not organized, which may include images, videos, PDFs, and text documents(Gandomi & Haider, 2015; Minelli et al., 2013).

Velocity signifies the increasing speed of data generation and transfer (Chen and Zhang 2014). Due to the development of mobile devices and other devices with sensors connected to the Internet, as well as the many types of streamed data from various sources, the contents of the data are continually changing. From this perspective, new algorithms and techniques are required to process and analyse the streaming and online data properly.

Veracity implies the accuracy and truthfulness of the data. It is unlikely to have all the data entirely correct and clean when dealing with a high volume, velocity and variety. The quality of the generated data can differ significantly so that the data accuracy of analysis is associated with the data source's veracity. Furthermore, it is vital to consider the reliability of data. For instance, a considerable amount of user reviews on social media tends to be fake, created by robotic sources. Given these facts, another aspect of big data is to deal with data integrity and accuracy to use it confidently (Affendey & Mamat, 2015; Saha & Srivastava, 2014).

Value is one of the most critical aspects of big data (Bello-Orgaz et al., 2016). The main aim is to discover substantial hidden values from large data sets. As such, big data is useless unless that value is revealed (Ghani et al., 2019).

2.8.2 Social Media Analytics

The contents of the web, from publisher to user-created content, have evolved due to the internet and Web 2.0 technologies' enhancements. It is now easily possible to access information such as topics, trends, sentiments, and reviews in a global manner from the internet. Furthermore, the dissemination and use of social media have given rise to immense opportunities and challenges for practitioners and researchers (Holsapple et al., 2018). According to the Digital 2020 October Global Statshot Report, more than 4 billion people worldwide now use social media each month, as well an average 2 million new users joining them. Social media usage by billions causes the generation of overwhelming data, and social media has turned to one of the most relevant and representative data sources for big data, which is widely regarded as “social big data” (Bello-Orgaz et al., 2016; Stieglitz et al., 2018). Big data creation via social media has occurred with the platforms such as social networking sites, blogs, microblogs, media sharing, wikis, question and answer sites, and reviews sites, which involve the important pieces of information concerning user interactions, daily activities, and sharing (Holsapple et al., 2018).

Zeng et al. (2010) define the term social media analytics (SMA) as “an emerging interdisciplinary research field that aims at combining, extending, and adapting methods for the analysis of social media data.” SMA involves big data analytics techniques such as machine learning, deep learning, text mining, and natural language processing. Fan and Gordon (2014) show the process of SMA consists of three main stages: capture, understand, and present. The capture stage refers to the identification of conversations on social media platforms with regards to activities and interests. This process is conducted by gathering a large amount of pertinent data from social media sources using application program interfaces (APIs), crawling techniques, and news feeds. The collected data are then prepared to create a dataset for the understanding stage. The data preparation implies several pre-processing techniques, including cleaning, data modelling, stemming, part-of-speech tagging, feature extraction, and other syntactic and semantic operations such as parsing. The understanding stage where advanced techniques such as text mining, machine learning, deep learning, and NLP stands at the core of SMA, as this stage must bring out actionable insight and uncover hidden patterns from the dataset. The most common methods of SMA, including sentiment analysis, social network analytics, text classification, opinion mining, and trend analysis, are implemented in this stage. Lastly, the present stage stands for the visualization of social media data. To make sense of a huge volume of data, visual analytics underpins data exploration and complex reasoning.

2.8.2.1 Social Media Analytics Techniques

2.8.2.1.1 Social Network Analytics

Social network analytics is concentrated on the investigation of the structural attributes of social networks, extraction of intelligence from the preferences and relationships amongst the entities (Scott & Carrington, 2011). The representation of the connected networks is provided using graphs, which are structures modelled through a set of nodes and a set of edges. The graphs are used to represent the extracted information where the nodes stand for participants while the edges represent relationships (Gandomi & Haider, 2015). Two of the most common graphs are social graphs and activity graphs (Heidemann et al., 2012). The social graphs structure comprises the binary and static social links (an edge between a pair of nodes that merely indicates the link's existence) among entities such as friendship or relationship, regardless of the actual activities of the entities. These graphs are mostly used for identifying communities or determining so-called hubs (i.e., users having a huge number of social links to others). Notwithstanding, users' actual communication activities are considerably important. Research points out that the value of social networking sites is measured by the communication activities among users. The network based on users' interaction is called an “activity network,” and the graph that stems from such a network refers to as an activity graph. In activity graphs, nodes symbolize users, while direct or undirected edges indicate activity links among users (Heidemann et al., 2012).

Several techniques capture information from social networks. A well-known technique is community detection which aims to obtain communities within a network (Fortunato, 2010). In a community graph, edges have global and nonuniform local distribution with strong connections of edges within special groups of nodes and relatively weak connections between them. In the same vein, communities are groups of nodes that potentially share common properties and/or play similar roles. The size of social networking sites tends to be tremendous, involving millions of nodes and edges. Summarizing such huge networks to extract behavioural patterns of users is at the core of the community detection techniques. In this respect, community detection is also regarded as clustering, and such problems can be solved with various clustering techniques (Fortunato, 2010). Community detection is used for many purposes in different domains. For instance, in marketing, clustering customers whose interests have similarity in a network of customers and products provide effective recommendation systems (Gasparetti et al., 2018).

Another example is that detecting the clusters of web clients whose interests have similarity and whose locations are close to each other tends to enhance services’ performances on the World Wide Web because a dedicated mirror server could be assigned to each cluster of clients (Krishnamurthy & Wang, 2000). Another common technique is social influence analysis, which indicates techniques regarding modelling and assessing the influence of actors and connections in a social network (Li et al., 2018). In a social network, an actor’s behaviour can be affected by others as a matter of course. For this reason, it might be considerable to measure the influence of participants, evaluate the strength of connections, and bring influence diffusion patterns into the open. In this respect, social influence analysis uses a variety of techniques pertinent to the mathematical structure of graphs to evaluate the significance of the network nodes. Such techniques include degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality (Li et al., 2018). Social influence analysis is a significant social indicator and is conducted in several fields such as viral marketing (Leskovec et al., 2005), health care communities (Fowler & Christakis, 2008), rumour spreading (Zaobo et al., 2015), and expert findings (Franks et al., 2014). To conclude, social influence analysis enables us to discover the social behaviours of users, facilitate theoretical support for decision-making and influencing public opinion, and underpin exchanges and propagation of diverse activities (Li et al., 2018).

2.8.2.1.2 Sentiment Analysis

Sentiment analysis is referred to automatically analysing opinions and sentiments from user-generated content, which is one of the most popular areas in NLP and widely used in data mining, web mining, and social media analytics (Deng & Liu, 2018). With the fast development in social media sites such as Twitter, Facebook, YouTube, and TripAdvisor, sentiment analysis draws considerable attention for both the industry and academia to extract key patterns of human thoughts.

Sentiment analysis tasks typically involve three different approaches: document-level sentiment analysis, sentence-level sentiment analysis, and aspect-level sentiment analysis. Document-level sentiment analysis tries to figure out the whole document's sentiment label with the assumption of having a single entity in document sentiment (Moraes et al., 2013; Tang, 2015). The sentiment labels could be positive and negative, thumbs up and thumbs down, or multiple categories such as review ratings 1 to 5 stars. Sentence-level sentiment analysis tasks aim to classify sentiment polarities in the given sentence (Jagtap & Pawar, 2013; Tan et al., 2011). Generally, a sentence's polarities are divided into three categories as positive, negative, and neutral. The task is regarded as a representative sentence classification problem. As such, sentence-level sentiment analysis tends to be more complicated than document-level techniques.

However, aspect-level sentiment analysis tasks attempt to classify sentiment polarities for an aspect in the given text unit. This technique identifies all sentiments within a text unit and detects the entity's aspects that each sentiment refers to (Schouten & Frasincar, 2015; Wang et al., 2016). User comments for a product, hotel, movie, and so on usually include opinions about different aspects. For example, a user review about a smartphone might contain different aspects of the products such as its battery, price, look, and technique features. Using aspect-level techniques allows us to attain information about different features of the product or service that could be missed if the sentiment is merely classified based on the polarity.

2.8.2.1.3 Text Mining and Natural Language Processing

"Text mining" refers to discovering and extracting information from unstructured text (Inzalkar & Sharma, 2015). As mentioned before, approximately 80% of data in the world is in an unstructured format. Thus, the massive increase in the amount of digital textual data has given rise to the generation of new insights and opportunities for research through new channels. Considering recent breakthroughs in natural language processing (NLP), machine learning, deep learning, and big data, text mining is now a more valuable and advantageous method for combining linguistic theories with NLP applications aiming to convert data from an unstructured format to a structured format (Gamallo & Garcia, 2019).

NLP, which stemmed from computational linguistics, make use of methods from diverse fields including computer science, artificial intelligence, and linguistics to allow computers to clutch human language in text or speech format (Chowdhary, 2020). Text mining and NLP have a bunch of overlapping and discrete strategies to work on unstructured text. NLP techniques aim to extract a complete meaning representation from the free text. This can roughly be defined as untangling the semantic meaning conveyed in the text, such as who did what to whom, why, where, when, and how (Kao & Poteet, 2007). The core of NLP is to use linguistic concepts, for example, part-of-speech (PoS) tagging, lemmatization, and anaphora resolution. Meanwhile, text mining techniques do not have to understand fuller meaning representation; their aim is the extraction of items of knowledge or regular patterns across several documents (Gamallo & Garcia, 2019).

Both hidden and new patterns can be extracted with the power of text mining and NLP relying on machine learning and deep learning in the big data environment. There exist several tasks concerning text mining and NLP, and the following part touches on some of the most common tasks containing information extraction, text summarization, text classification, sentiment analysis, and question answering.

2.8.2.1.4 Information Extraction

Information extraction (IE) focuses on extracting structured data such as entities, the relationship between entities, and attributes describing entities from the unstructured or semi-structured text (Hobbs & Riloff, 2010). IE's three common sub-tasks with NLP are part of speech (PoS) tagging, named entity recognition (NER), and relation detection. Part of speech (PoS) tagging, also known as grammatical tagging, is the technique of identifying the part of speech for a particular word or portion of the text according to its usage (Gimpel et al., 2010). For example, the words in the given sentence "London is a beautiful city in England" are tagged with NLTK, a natural language processing toolkit, based on their context as [("London," "NNP"), ("is," "VBZ"), ("a," "DT"), ("beautiful," "JJ"), ("city," "NN"), ("in," "IN"), ("England," "NNP")] as seen in Figure 4. In the example, the tags NNP, VBZ, DT, JJ, NN, and IN represent singular proper noun, verb, determiner, adjective, and singular noun, respectively. In addition, PoS can detect homonyms for word-category disambiguation. For instance, in the sentence "They refuse to permit us to obtain the refuse permit," the words refuse and permit show up twice as a verb and a noun with different meanings. PoS tagging here detects what words are verb and what words are noun [(They, PRP), (refuse, VBP), (to, TO), (permit, VB), (us, PRP), (to, O), (obtain, VB), (the, DT), (refuse, NN), and (permit, NN)].

Figure 4 A Simple Part of Speech Tagging with NLTK

```
text=nltk.word_tokenize('London is a beautiful city in England')
tagged=nltk.pos_tag(text)
print(tagged,end='')
```

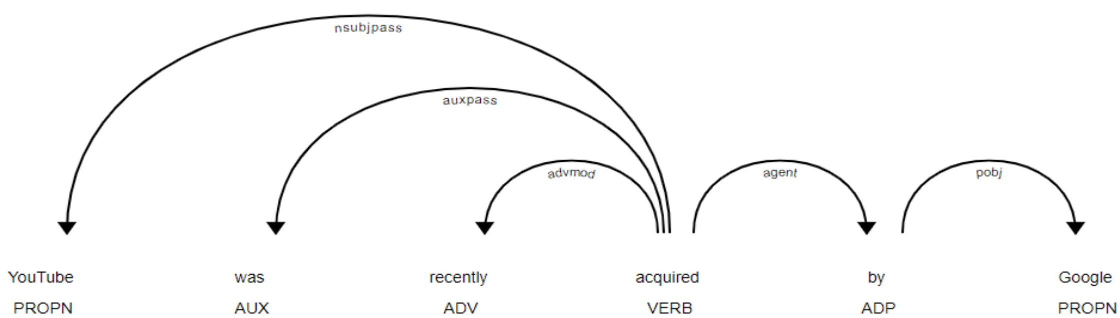
```
[('London', 'NNP'), ('is', 'VBZ'), ('a', 'DT'), ('beautiful', 'JJ'), ('city', 'NN'), ('in', 'IN'), ('England', 'NNP')]
```

Named entity recognition (NER), which is also defined as entity identification or extraction, tasks identify names of people, places, organizations, dates, locations, and other entities of interests in text (Li et al., 2020). Take the sentence "George worked at Google in California in 2010" as an example. NER allows us to see the types of entities in the sentence, as seen in Figure 5**.**

Figure 5 Simple Named Entity Recognition With Spacy

Relation extraction (RE) looks for relations between specified types of named identity (Pawar et al., 2017). A way of performing this technique is to seek for all triples of the form (A, α, and B) in which A and B are named entities, and α is a set of words between X and Y. Subtree matching is an important indicator to find relations as it shows dependency paths between entities. For illustration, given the sentence "YouTube was recently acquired by Google," the dependency between the entities is shown as in Figure 6, which is implemented with the Spacy tool. Here, the dependency tag for YouTube is nsubjpass that refers to a passive subject. The other entity, "Google," is the object in the text and the term "acquired" is the ROOT of the sentence, meaning that it connects the object and subject.

Figure 6 Finding Dependency by Spacy



```
for tok in doc:
    print(tok.text,"-->",tok.dep_,"-->",tok.pos_)
```

```
YouTube --> nsubjpass --> PROPN
was --> auxpass --> AUX
recently --> advmod --> ADV
acquired --> ROOT --> VERB
by --> agent --> ADP
Google --> pobj --> PROPN
```

2.8.2.1.5 Text Summarization

Text summarization focuses on generating a summary of one or more text documents preserving the key points and the original document's content meaning. The explosion in the amount of data from a wide range of sources in the big data era has led to a need for developing machine learning and deep learning algorithms in the NLP field, providing automatic summarization for the purpose of shortening long documents and delivering appropriate summaries. Current methods to automatic summarization can generally be divided into two categories: extraction and abstraction (Allahyari et al., 2017; Gambhir & Gupta, 2016).

Extractive summarizations run by picking up key phrases or sentences in the original document to form a consistent summary. Extractive summarizers typically carry out the following tasks: creating an intermediate representation of the input text, scoring the sentences based on the representation, and generating a summary by picking the most important sentences (Nenkova & McKeown, 2011). By contrast, abstractive summarizations initially construct an internal semantic representation and then perform natural language generation techniques to produce a summary. The summaries might convey the text units that are not compulsorily present in the original document (Gambhir & Gupta, 2016). Compared with extractive methods, abstractive methods are more difficult to adopt (Gudivada et al., 2015). Advanced abilities that are significant to high-level summarization such as paraphrasing and generalization can be provided only in abstractive methods, thanks to the advanced NLP, and deep learning techniques.

2.8.2.1.6 Text Classification

Text classification is one of the ultimate tasks in text mining and NLP to assign labels to text. It includes a wide range of applications from topic modelling, sentiment classification, and spam detection. In general, the text classification system consists of four different levels of scope and can be selected based on the objective of the task (Kowsari et al., 2019), as show in the following:

- **Document level** acquires the pertinent categories of the full document
- **Paragraph level** acquires the pertinent categories of a single paragraph
- **Sentence level** acquires the pertinent categories of a single sentence
- **Sub-sentence level** acquires the pertinent categories of sub-expressions within a sentence

Document-level classification is appropriate for the tasks when the analysis of the complete meaning of the document needed. For instance, to classify documents into topics such as technology, entertainment, politics, the document-level approach is suitable as using whole document provides more information and context to the algorithm. The paragraph-level approach is quite useful if different information or categories exists in different paragraphs, so analysing the document in paragraph level helps extracting information for each paragraph (Ferguson et al., 2009). However, splitting the document into smaller chunks could be beneficial to obtain granular results (Zirn et al., 2011). For instance, a customer review may include both positive and negative comments regarding a product. In this case, splitting text into smaller units allows us to understand what exactly the customer likes or dislikes. Given similar scenarios, using sentence level and sub-sentence level classification approaches can provide better results.

Apart from the abovementioned rules, it is noteworthy to mention different text classification types based on the number of categories or labels assigned to the text. Here, we can categorize the classification tasks into three categories: binary text classification, multiclass text classification, and multilabel classification.

Binary text classification is the most simplistic text classification task in which each item falls into one of the two categories exclusively (Jo, 2018). The goal of detecting an email as spam or not is a binary classification task example. Another example is identifying a text's sentiment based on positivity or negativity. Multiclass text classification is comprised of more than two classes where labels are mutually exclusive. The assumption underlying multiclass text classification is that each sample has to be assigned to one and only one label (Gargiulo et al., 2019). However, multilabel text classification assigns a label or set of labels to each sample, with labels that are not mutually exclusive. Thus, there is no limitation on the number of categories the samples can be assigned to (Azarbonyad & Marx, 2019; Gargiulo et al., 2019). For instance, a toxic comments dataset, which is available at Kaggle, is suitable for the multilabel classification task. The comments in the dataset include different types such as threats, insults, or obscenities. Each comment can include more than one toxicity type or none of them, and the aim of multilabel text classification is to assign label or labels to each comment based on their content.

The use of deep learning in NLP has remarkably increased the performance of text classification tasks where conventional machine learning models for NLP often did not have the ability to absorb a large amount of training data. Some of the most common and efficient models for binary or multiclass text classifications are convolutional neural networks (CNN) (Kalchbrenner et al., 2014), recurrent neural networks (RNN) based on long short-term memory (LSTM) (Zhou et al., 2015), and hierarchical attention network (HAN) (Yang et al., 2016). Bidirectional Encoder Representations from transformers (BERT) (Devlin et al., 2018) has also been a common and efficient model for multilabel text classification tasks.

2.8.3 Advanced Analytics

2.8.3.1 Machine Learning

Machine learning is a vast research area based on using computational methods to find usable patterns in empirical data (Edgar & Manz, 2017). In other words, it stands for training machines using available data and making it possible to produce models that could analyse larger and more complicated data to make more accurate results for the given tasks

(Mohri et al., 2018). Machine learning is at the core of big data analytics with its abilities to extract actionable insights from large datasets. Techniques based on machine learning have been implemented in various domains ranging from education, engineering, finance, and medicine (El Naqa & Murphy, 2015). Marketing is also a field taking advantage of machine learning on a large scale. The proliferation of ubiquitous interactions between customers and firms has led to the generation of a vast amount of data, which has impelled companies to invest in machine learning technologies to improve marketing capabilities. Based on BCC research, the growth in machine learning-enabled solutions will be at a 43.6% annual rate, equal to $8.8 billion by 2022 in the global market (Ma & Sun, 2020). State-of-the-art machine learning algorithms have been performed in recommendations systems, content optimization, image recognition, speech recognition, natural language processing, social media mining, and many other applications. These breakthroughs have all enhanced business performance in many aspects.

The following part will include an overview of machine learning tasks.

2.8.3.1.1 Supervised Learning

Supervised learning is based on making predictions and classifications with the existing evidence to resolve uncertainty. The input data set is split into a training dataset, a known set of input data, and a test dataset as output, a known response to the data, where the supervised learning model performs to learn from the training dataset to produce rational predictions as for the response to the new data. Given that, supervised learning is quite suitable if the data are known for the output trying targeted for prediction (Caruana & Niculescu-Mizil, 2006).

Supervised learning can be used for classification and regression problems to build predictive models. Classification techniques use a model to correctly allocate a category to each item, technically assigning test data into discrete categories. Classification identifies specific entities in the dataset and tries to make inferences regarding how those entities needed to be labelled or categorized. Common classification algorithms are Naïve Bayes, Support Vector Machines, Logistic Regression, Decision Trees, *k*-nearest neighbour, and Neural Networks. Moreover, regression techniques are used to figure out the relationship between dependent and independent variable(s) and make predicting a real value for each item. Typical examples of regression contain stock forecast, estimation of house prices, and sales revenue of a business. Popular regression algorithms consist of linear regression, nonlinear regression, and regularization. One of the crucial points in deciding between classification and regression techniques is the nature of the data. Classification techniques are appropriate if the data can be categorized, labelled, or split into defined classes while regression techniques are useful if a data range exists, or the nature of the output is numerical. In other words, the key distinction between classification and regression is that classification predicts labels or classes, and regression predicts a continuous value or quantity (Ma & Sun, 2020; Mohri et al., 2018; Talabis et al., 2014).

2.8.3.1.2 Unsupervised Learning

In unsupervised learning tasks, the objective is to reveal hidden patterns or extract intrinsic structure in data without needed human intervention in which the training set includes only the input variables when the output variables are either unlabelled or unknown (Talabis et al., 2014). The most common task of unsupervised learning is clustering analysis. Clustering is a technique used to group unlabelled data or find hidden patterns to maximize similarities within-groups or differences between groups. The algorithms belonging to clustering analysis can be classified into a few types: k-Means clustering, hierarchical clustering, and probabilistic clustering (Stone et al., 2017)**.** Dimensionality reduction is a task used to transform data from high dimensional to lower dimensional with keeping the integrity of the dataset as far as possible. As generally accepted, more data can lead to more accurate results, having high-dimensional data could be undesired for some reasons. As such, high-dimensional data can computationally be difficult when analysing and impact the performance of machine learning algorithms. Some of the most common applications of this technique are data visualization, noise reduction, and clustering analysis. Dimensionality reduction techniques include but not limited to principal component analysis (PCA), singular value decomposition (SVD), autoencoders, and nonnegative matrix factorization (NMF) (Van Der Maaten et al., 2009). An association rule is a rule-based method seeking to discover remarkable hidden relationships between variables in the given large datasets. Association rule methods help the items that frequently appear together and are usually used in market-based analysis, allowing companies to better perceive relationships between various products and what items customers generally purchase together. Thus, consumption habits can be analysed, and better selling strategies, as well as recommendation engines, could be developed (Borek et al., 2013).

2.8.3.1.3 Semi-Supervised Learning

Semi-supervised learning can be seen as a mixture of supervised learning and unsupervised learning. In fact, it is necessary to have a large amount of labelled data for supervising learning methods. However, data labelling or annotating can typically be carried out at a considerable cost because it usually requires experts to get it conducted manually. In contrast, semi-supervised learning often involves much fewer labelled samples than unlabelled samples. Semi-supervised learning allows algorithms to learn from the small amount of labelled data using pseudo labelling, including several neural network models, and training methods. A number of semi-supervised learning methods include graph-based methods, heuristic approaches, and low-density separation (Mohri et al., 2018; Zhu & Goldberg, 2009).

Considering this era where the data are growing exponentially, it is pretty difficult to label each data analysis scenario. A wide range of problems arising in applications, including classifications, regressions, and clustering can be adapted as instances of semi-supervised learning, leading to better performance with the distribution of accessible unlabelled data.

2.8.3.1.4 Transfer Learning

Transfer learning is a deep learning approach where a pretrained model can serve as a starting point to train a model for a different purpose or dataset in the given task. Transfer learning is an efficient technique for models with a small amount of labelled data because it can dramatically reduce computation time in large datasets. The models with transfer learning do not need to be trained as many times as a new model would (Pan & Yang, 2009).

2.8.3.1.5 Active Learning

Active learning tasks are techniques in prioritizing data needed to be labelled to have the strongest influence in training the model. Active learning can be useful when the data are too big to be labelled, and some priority is needed to label the data efficiently. The objective is to maximize prediction performance by minimizing the data requirement (Settles, 2009).

2.8.3.1.6 Reinforcement Learning

Reinforcement learning is a general computational approach of learning from the action without a network trainer or a supervisor through trial-and-error experience to figure out which actions provide the greatest reward. Reinforcement learning contains three principal components: the agent, the environment, and the actions. The agent continually interacts with the environment to learn and receive rewards from taking actions. This learning is objective or task-based; the agent learns how to achieve its objective by performing the best possible actions to maximize the reward for the predefined time. Reinforcement learning offers autonomous systems to learn from their experience and is often used for gaming, robotics, and morphing websites (Sutton & Barto, 2018; Wiering & Van Otterlo, 2012).

2.8.3.2 Deep Learning

The principal notion of deep learning is to automate feature extraction and pattern identification from complicated unsupervised data without human intervention, thereby playing a vitally significant role in big data analytics (Bengio et al., 2013). In deep learning, supervised or unsupervised techniques are performed to automatically learn and extract complex data representations (Jan et al., 2019). These algorithms have been greatly inspired by artificial intelligence with biological observations on human brain mechanisms, and their main purpose is to emulate the hierarchical learning approach of human brains to analyse, learn, and make decisions for excessively complicated tasks. In contrast to traditional shallow machine learning algorithms, which have limitations in dealing with unstructured and complex problems because of their simple architecture, like support vector machines, Naïve Bayes, and decision trees; deep learning techniques have the ability to extract high-level features and to learn the hierarchical representations by merging the low-level input data from complicated structures and relationships in the input corpus (Najafabadi et al., 2015; Zhang et al., 2018).

Deep learning algorithms are composed of deep architectures of sequential multilayers. Each layer has an input and an output. After a nonlinear transformation applied in its input, a representation is provided in its output. The goal here is to detect complicated features hierarchically by transiting the data throughout multiple transformation layers, with simple features being identified by the lower layers and fed into higher layers. Feeding the nonlinear transformational layers is the fundamental concept in deep learning. The number of layers the data passes through is proportional to the complication of nonlinear transformation. Deep learning tasks learn the representation of the data by these transformations because deep learning algorithms can be regarded as a special form of representation learning algorithms (Najafabadi et al., 2015).

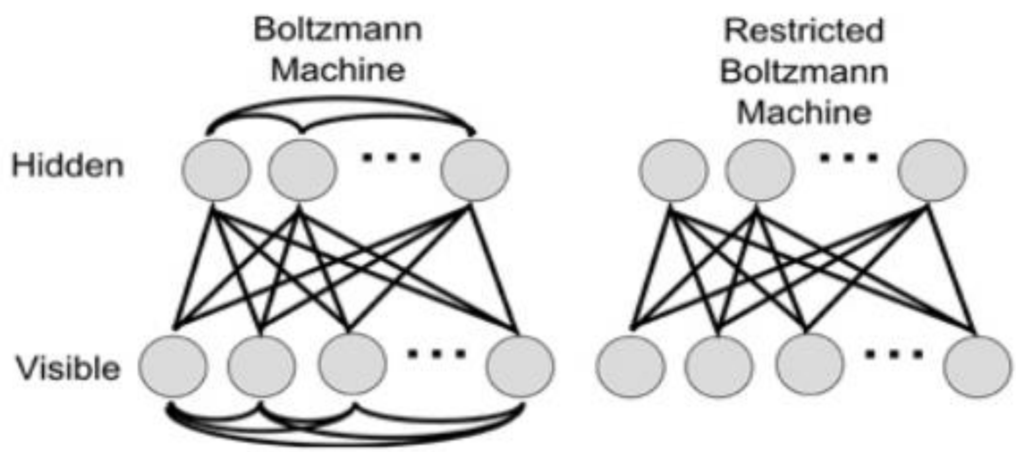
Learning the parameters in deep learning algorithms work through a greedy layer-wise unsupervised pretraining, as proposed by Hinton (2006). Each layer is trained individually, and the output of the former layer is provided as the input layer (learning layer) for the next layer. The iteration carries on until the desired number of layers is acquired, and then the training phase of deep learning completes. At this point, layers with learned weights can be performed in a supervised fine-tuning stage as initialization for other tasks such as classification and recognition.

There are two main building blocks that unsupervised single layer learning algorithms used to generate deeper models (Najafabadi et al., 2015). autoencoders (AEs) and restricted Boltzmann machines (RBMs):

- AEs are networks composed of three layers which are input, hidden, and output. The concept in AEs is trying to grasp some representations of the input in the hidden layer to rebuild the input in the output layer depending upon these mediate representations, and thence the target output is the input itself (Baldi, 2012). Back-propagation algorithm is used to train AEs.

- RBMs are considered as the most popular type of Boltzmann machine. As seen in Figure 7, they consist of one visible and one hidden layer with the condition of the nonexistence of interactions between the units in the same layer units, and the existence of interactions between units in discrete layers (Jan et al., 2019). RBMs are probabilistic generative models that can learn data distributions without data label information. Therefore, they could manage huge, massive amounts of unlabelled data for controlling complicated data structures. The Contrastive Divergence algorithm is mostly performed to train BMs (Xue-Wen & Xiaotong, 2014).

Figure 7 Boltzmann Machine and Restricted Boltzmann Machine



Source: O’Connor et al (2013).

A vast number of deep learning models have been built so far. The most typical deep learning models contain deep belief network (DBN), and convolutional neural network (CNN) are mentioned as follows.

2.8.3.2.1 Deep Belief Networks (DBN)

A deep belief network (DBN) has a deep architecture with the ability to capture feature representations from both labelled and unlabelled data. The model built by DBN includes the combination of unsupervised pretraining and supervised fine-tuning approaches;

unsupervised phases try to learn data distributions without using data labels while supervised phases commit local search for fine-tuning (Hinton, 2009).

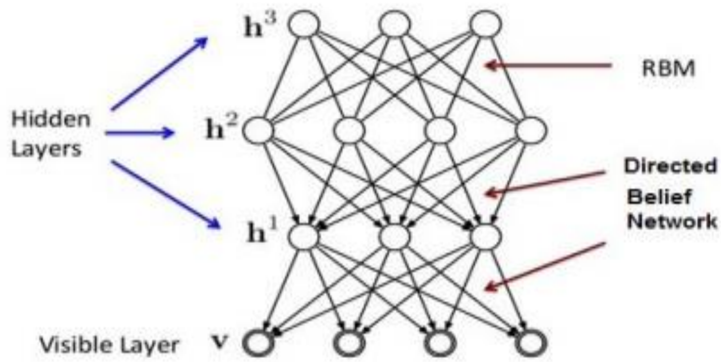
A DBN architecture, as shown in Figure 8, comprises the input layer, hidden layer, and output layer. It comprises a stack of RBMs as there are connections between each two adjoint layers while there is no connection between the units in the same layer. Pre-training of RBMs is processed layer-by-layer which is crucial because it helps overfit issues that can occur when millions of parameters are used (Xue-Wen & Xiaotong, 2014).

DBNs learning stage can be illustrated as follows (Elaraby et al., 2016):

- The first layer training is performed as an RBM, shaping the raw input as its visible layer.
- The second layer uses the input data representation (this representation can be received as being either sample of conditional probabilities or mean activations) acquired by the first layer as training data.
- Then the first layer turns into a visible layer to the second layer. Besides, training of the second layer is performed by receiving the transformed data (mean values or samples) as training instances.
- Steps 2 and 3 are repeated for each layer; transformed data are propagated each time upwards.

The initialization of all weights of the layers is now completed. Supervised fine-tuning for the entire network can be carried by adding a final layer to represent the desired outputs and derivatives of back-propagating for better discrimination.

Figure 8 The Deep Belief Network as a Stack of RBMs



Source: Hinton et al., (2006).

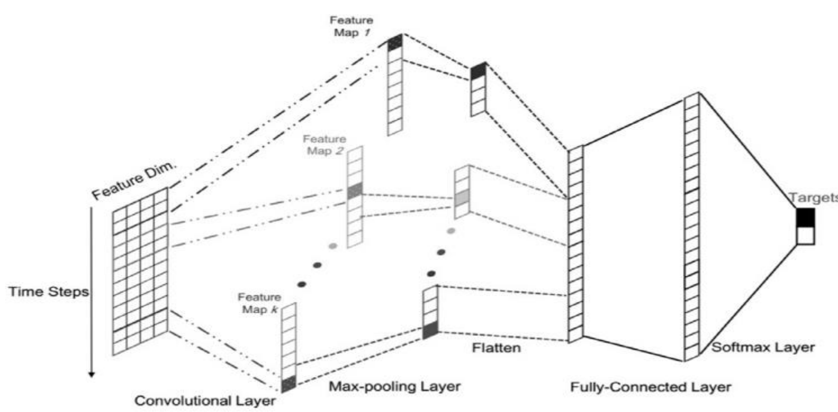
2..3.1.2 Convolutional Neural Networks (CNN)

The convolutional neural network is originally designed for image analysis. However, it has been witnessed that CNN also has an excellent ability in sequential data such as natural language processing (Albawi et al., 2017). A typical convolutional neural network is comprised of numerous hierarchical layers, including feature map layers and classification layers, as shown in Figure 9. It frequently commences with a convolutional layer and subsampling layer (pooling layer); the convolutional layer accepts data from the input layer and is in charge of convolution operations with many same size filter maps to achieve weight sharing while the subsampling layer is responsible for reducing the size of proceeding layers (Xue-Wen & Xiaotong, 2014).

To exemplify, we can take a 2D image x. The image is initially dissociated into a consecutive input as $x = \{x_1, x_2, x_3, \dots, x_N\}$. The convolutional layer is defined as the nonlinear activation function as: $y_j = f\left(\sum_i K_{ij} \otimes x_i + b_j\right)$, in which y_j stands for jth output for the convolutional layer, K_{ij} stands for the convolutional kernel (or trainable filter) that convolves with the feature map x_i from prior layer to generate a new feature map in the current layer, $\otimes$ stands for the discrete convolution operator, b_j stands for the bias, f is a nonlinear function (mostly a scaled hyperbolic tangent function). The subsampling layer aims to reduce the feature map's dimension, which can be performed by a max pooling operation over a small region or averaging $2 * 2$ areas in the feature map. Following, fully connected layers and as the softmax layer are mostly placed on the top layer for recognition and classification (Xue-Wen & Xiaotong, 2014; Zhang et al., 2018).

In CNN models, defining the weights between layers is crucial, which are often trained with standard back propagation techniques and a gradient descent algorithm including the mean squared error as the loss function. However, CNN architectures could be trained with unsupervised techniques such as predictive sparse decomposition (PSD). Briefly, CNN models capture a hierarchical feature representation by performing techniques such as shared weights (using the same weights to build the feature maps tends to decrease the number of parameters remarkably) and subsampling (to reduce the dimension of the feature map). A CNN model has the capacity to automatically learn high-level features and perform well, especially in image analysis and natural language processing with the recent developments (Xue-Wen & Xiaotong, 2014).

Figure 9 Convolutional Neural Network



Source: (Zhao et al., 2019).

CHAPTER 3:

METHODOLOGY

3.1 Introduction

This chapter covers different research methodologies to address the most convenient. Within this context, various research philosophies and approaches, along with the justifications of methodological choices and adopted research methods, are introduced.

Three research philosophies (positivism, interpretivism, and pragmatism) are described in Section 3.2 to adopt the most convenient philosophy for this research. Next, Section 3.3 covers an overview of the research design and elucidates the general research plan in a flow chart. Section 3.4 then focuses on three research approaches: deductive approach, inductive approach, and abductive approach, explaining the rationale of the adopted approach. Finally, the chapter concludes with Section 3.6.

3.2 Research Philosophy

Collis and Hussey (2013) describe a research philosophy as the philosophical view and assumptions of a researcher with respect to how research should be carried out to build and embrace knowledge. In the light of this statement, it is highly significant for this research to comprehend its nature to investigate useful and the most convenient methods. According to Saunders et al. (2019), assumptions will contribute to the selection of a research method and strategy. They define that there are three major approaches to research philosophy: ontology, epistemology, and axiology.

3.2.1 Ontology

Ontology concerns the nature of reality, which implies the researchers' assumptions about how the world functions and their adherence to views. Objectivism and subjectivism are the two ontological aspects. *Objectivism* maintains that social reality is external to us and others (social actors). Objectivism accepts realism, which takes social entities to be like physical entities of the natural world, in that they exist regardless of how we think of them, name them, or even be aware of them. Because social actors' interpretations and experiences do not affect the social world, an extreme objectivist believes all experience only one actual social reality (Saunders et al., 2019). This social universe consists of substantial, granular, and relatively unchanging 'things,' such as family, religion, and the economy (Burrell & Morgan, 2017). Objectivists believe social and physical phenomena exist irrespective of individual ideas and are universal and permanent. *Subjectivism* assumes social reality is based on the perceptions and acts of social actors (people). Ontologically, subjectivism supports nominalism. Extreme nominalism holds that the order and structures of social phenomena (and the phenomena themselves) are constructed by researchers and other social actors through language, conceptual categories, perceptions, and actions. There is no reality beyond what individuals attach to the social world (Saunders et al., 2019). Because each person experiences and perceives reality differently, it makes more sense to speak of various realities as opposed to a single reality shared by all (Burrell & Morgan, 2017). Less extreme is social constructionism, which advocates that reality is produced by social interaction where social actors establish partially shared meanings and realities. Alternatively put, reality occurs intersubjectively. As social interactions between actors are ongoing, social phenomena are in flux and change. This means a researcher must examine a situation's historical, geographical, and sociocultural circumstances to understand what is happening or how realities are experienced. A subjectivist researcher is interested in varied viewpoints and narratives that can account for different social realities of different social actors, unlike an objectivist researcher who seeks universal facts and laws regulating social behaviour (Saunders et al., 2019).

3.2.2 Epistemology

Epistemology refers to assumptions about what constitutes acceptable, valid, and legitimate knowledge and how knowledge can be communicated (Burrell & Morgan, 2017). The multidisciplinary nature of business and management allows for the legitimacy of several sorts of knowledge, ranging from numerical data to textual and visual data, facts to views, and tales and stories. Consequently, different business and management academics employ diverse epistemologies in their studies, such as projects based on archival research and autobiographical stories (Marti & Fernandez, 2013), narratives (Gabriel et al., 2013), and fictional literature (De Cock & Land, 2006). This diversity of epistemologies offers a wide selection of techniques. Nonetheless, it is essential to comprehend the consequences of various epistemological assumptions concerning the choice of method(s) and the strengths and limitations of the subsequent study findings. For instance, the (positivist) idea that objective facts provide the best scientific evidence is likely, but not always, associated with selecting quantitative research methods. As a result, the accompanying research findings will likely be considered objective and generalizable. However, they will be less likely to present a rich and nuanced vision of organizational realities, account for the disparities in individual circumstances and experiences, or possibly propose a completely new understanding of the world than if they were founded on a different view of knowledge. In other words, despite this variation, epistemological assumptions help determine the legitimacy of your research (Saunders et al., 2019).

The most common epistemological stances are positivism and interpretivism. Positivism is based on working with observable reality with the belief of society leading to the production of generalization, while interpretivism is associated with in-depth variables and factors related to a context, and it considers that individuals create further depth in meanings (Alharahsheh & Pius, 2020). Table 1 indicates the comparison of research assumptions between positivism and interpretivism.

Table 1 Assumptions of Positivism and Interpretivism

Assumptions	Positivism	Interpretivism
Ontology	Reality is objective and apart from the researcher	Reality is subjective and seen by participants in the study
Epistemology	The researcher is independent of that being researched	The researcher interacts with that being researched
Axiological	Value-free and unbiased	Value-laden and biased
Methodological	Deductive Process Cause and Effect Static design (categories isolated before the study) Context-free Accurate and reliable through validity and reliability	Inductive process Mutual simultaneous shaping of factors Emerging design (categories identified during the study) Context-bound Accurate and reliable through verification
Data Collection	Highly structured Large Samples Quantitative	In-depth investigations Small Samples Qualitative

Source: Adapted from Collis and Hussey (2003, p.49), and Saunders et al. (2009).

3.2.2.1 Positivism

In the positivist assumption, the reality can be described objectively using measurable properties and depends on the researcher. Accordingly, positivism is associated with statistical analysis and quantitative methods (Collis & Hussey, 2003; Saunders et al., 2019). Quantitative methods can explain and predict conditions in the social world by investigating underlying relationships between the components that exist (Collis & Hussey, 2003). Studies adopting a positivist approach make use of existing literature to build theories and establish hypotheses so that it is needed to gather numerical data for investigating and understanding human behaviours (Saunders et al., 2019). As the gathered data are numerical, the results generally suggest valid or invalid outcomes, more widely understood as acceptance or rejection of the hypotheses developed or tested (Saunders et al., 2019).

3.2.2.2 Interpretivism

Conversely, in the interpretivist approach, the reality cannot be measured objectively as it is in the researcher's mind as subjective. Moreover, the reality is influenced by how it will be examined (Collis & Hussey, 2013). Interpretivism is therefore linked with the inductive approach, qualitative methods, and in-depth analysis to comprehend social phenomenon. Besides, in this sort of research paradigm, the knowledge is usually hidden and requires to be revealed by deep analysis (Alharahsheh & Pius, 2020). Hence, for studies adopting interpretivist approach, the qualitative method is more proper to examine and understand human behaviours (Saunders et al., 2019).

3.2.2.3 Pragmatism

Apart from positivism and interpretivism, Saunders et al. (2019) highlight another perspective: pragmatism. The discussions around both epistemology and ontology have unavoidably sounded competitive. The decision between the positivist and interpretivist research philosophy is frequently used to frame the debate. However, choosing between one position and the other is somewhat unrealistic in practice from the perspective of pragmatism. A pragmatist paradigm involves that research begins with a problem and tries to find practical solutions that enlighten future practice. In a pragmatist approach, the most significant determinant is the research question for the design and strategy of the research. Saunders et al. (2019) point out that if a research question does not explicitly adopt a particular knowledge or method, it is utterly possible to combine different sorts of knowledge and methods. Because the problem lies behind this paradigm instead of methods (Patton, 2014), all approaches can be deployed to understand the problem. In this sense, pragmatists consider many ways of conducting the research rather than sticking to a single point of view.

3.2.3 Axiology

Axiology is a discipline of philosophy concerned with evaluating value judgments. Despite the fact that this may incorporate ideals we possess in the realms of aesthetics and ethics, the process of social inquiry is the focus here. If we want our research results to be credible, the role that our values play at every stage of the research process is of utmost importance

(Saunders et al., 2019). Heron (1996) asserts that our values are the guiding principle behind all human action. Moreover, he claims that researchers display axiological expertise by expressing their values as a basis for deciding the type of study they perform and how they conduct it. Ultimately, researchers will demonstrate the values at every phase of the research process.

3.2.4 Adopted Research Philosophy

This research has been conducted with the pragmatist approach that claims that concepts only matter in which they support action. It also strives to accommodate objectivism and subjectivism, fact and values, accurate and rigorous knowledge, and different contextualised experiences. With the pragmatism view, this research evaluates theories, concepts, ideas, and research findings not in an abstract manner but terms of their roles as instruments of thought and action and their practical outcomes in particular situations. As this research is more concerned with actual outcomes than abstract distinctions, the nature of this research does not strictly support just one aspect of ontology as 'objectivist' or 'subjectivist'. The reality is vital for this research, as the practical implications of ideas and the flow of processes and practices. Knowledge is also valued because it enables actions to be carried out successfully.

This research also supports that there are many ways of interpreting the world and undertaking research and that no single point of positivist or interpretivist view can ever give the entire picture. In line with pragmatism, the most crucial determinant for the design and strategy of this research is the research problem to be addressed and the research questions which incorporate the pragmatist emphasis on practical solutions that inform future practice. The research problem in this research does not suggest unambiguously that one particular type of knowledge or method should be adopted; this only confirms the pragmatist's view that it is perfectly possible to this research with different kinds of knowledge and methods. This means that multiple methods are often possible and possibly highly appropriate within one study. Accordingly, this research adopts the mixed-method consisting of qualitative and quantitative methods rather than having a single point of view.

Moreover, the role of values has unavoidable importance in pragmatist research philosophy (Dudovskiy, 2016). The research philosophy reflects the researcher's values, as is the choice of data collection techniques. In this research, social media has been valued as a data collection source to analyse a vast amount of data for a comprehensive analysis. In addition, values drive the reflexive inquiry process, initiated by doubt and a sense that something is lacking or out of place in the chosen research topic. That recreates belief when the problem has been resolved with the big data from social media analytics as a research strategy that the researcher values.

3.3 Research Method

In business and management research methods, quantitative and qualitative are frequently used to differentiate data gathering approaches and processing procedures. Concentrating on numeri or non-numerical data is one technique to distinguish between these two. Quantitative is commonly used as a synonym for any data gathering method (such as a questionnaire) or data analysis tool (such as graphs or statistics) that creates or uses numerical data. In contrast, qualitative is commonly used as a synonym for any data gathering technique (such as an interview) or data analysis method (such as classifying data) that generates or uses non-numerical data.

This research constitutes an exploration of contextual management response strategies to online reviews. For this, a qualitative method is adopted; a content analysis including data annotation is conducted over a sample, and then several models are created with deep learning-based text classification approaches to categorize a large amount of unstructured text data. The quantitative methods are usually employable when researchers aim to examine relationships between numerically measured variables that are obtained with well-structured data collection techniques (Saunders et al., 2019). The data are then transformed into structured data by quantifying the qualitative data. As such, textual data are represented numerically. In addition to that, various numerical variables consisting of response speed, response rate, response length, review valence, review volume, and review variance are used in the empirical analysis to examine the relationship between these characteristics and customer ratings.

Overall, this study adopts a mixed-methods approach, which is an integration of qualitative and quantitative research methods. Mixed-methods research combines these methods by collecting, analysing, and mixing both quantitative and qualitative data in single research (Saunders et al., 2019). Data obtained for this thesis consists of numeric and unstructured textual data, and the qualitative textual data are quantified and turned into quantitative instruments to analyse statistically further.

3.4 Research Strategy

The research strategy is seen as the methodological bridge between the opted research paradigm and the method preferences of researchers for the analysis and collection of data (Saunders et al., 2019). It also stands for the researcher's plan to implement prosperous research by sufficiently addressing the research questions (Bryman & Bell, 2015). Saunders et al. (2019) state that the research strategies are varied as experiment, survey, archival research, case study, ethnography, action research, grounded theory, and narrative inquiry. By looking at these research strategies, experiments and surveys are primarily associated with quantitative research design; archival research and case studies are mainly associated with mixed-method design (a combination of quantitative and qualitative techniques). The remaining research strategies are principally related to qualitative research design (Saunders et al., 2019).

This study adopts empirical research, employing big data and social media analytics approaches to a large set of secondary data collected from an online social networking platform. Social media analytics is an interdisciplinary research field aiming to combine, extend, and employ methods for the analysis of social media data. In this research, nearly 400,000 rows are collected and analysed using several computational techniques to present insights about organizational strategies.

3.5 Research Design

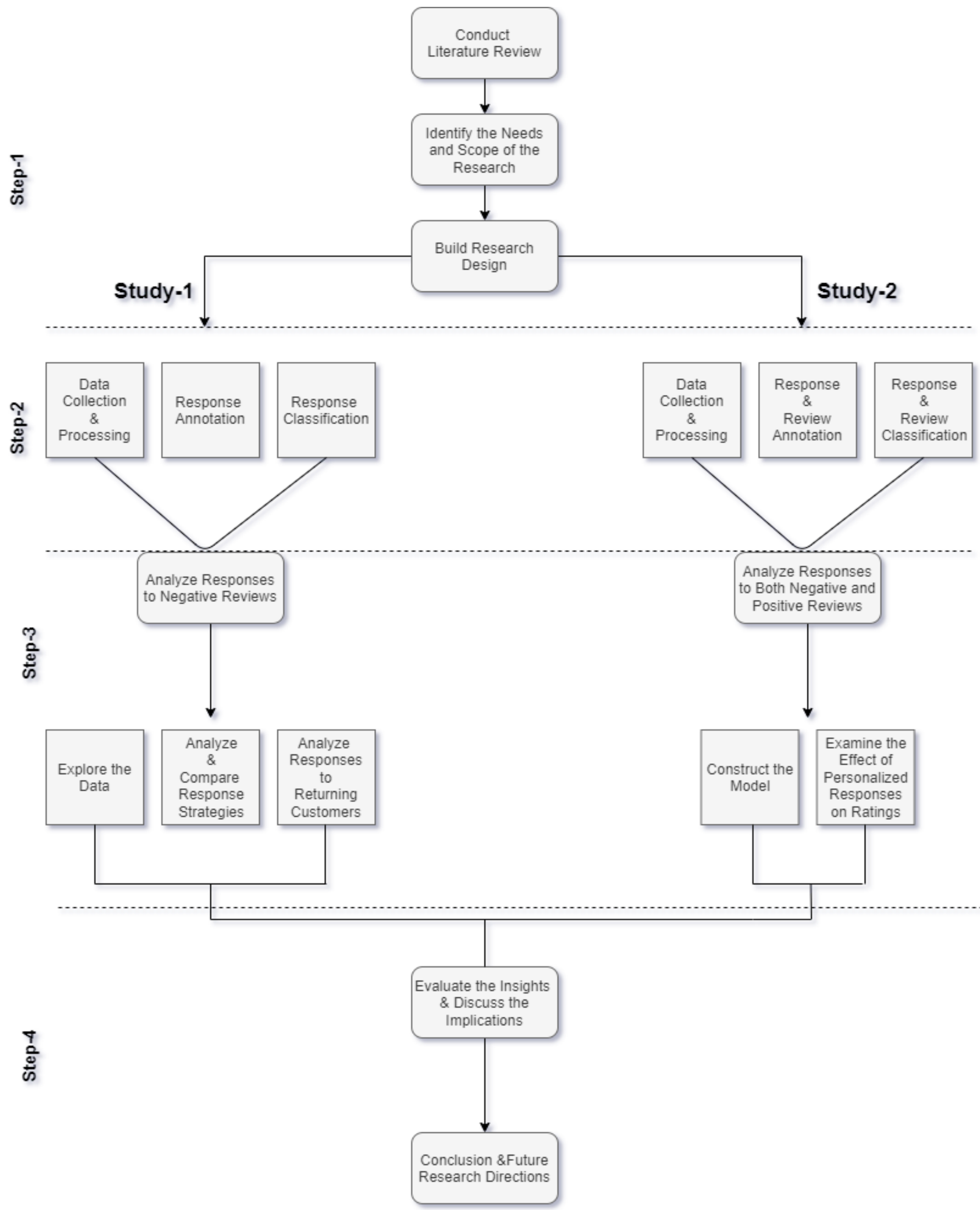
Research design is described as the overall research plan that assists in acquiring answers to research questions (Saunders et al., 2019). Building an elaborate plan that is inclusive of clear research objectives, consistent research questions, a well-structured data collection, and techniques for data analysis could provide an advantage for researchers to guide and concentrate on their studies (Saunders et al., 2019).

Figure 10 shows the research design implemented in four main steps for the two studies. In the first step, a comprehensive literature review of management responses and social media big data analytics is conducted. Subsequently, the scope and needs of the research are determined, and the research is designed for two independent studies. In the second step of Study–1, management responses to negative reviews are collected, processed, annotated, and classified. The third step of Study–1 consists of an exploratory analysis with descriptive statistics and numerous data visualizations that carried out to reveal what strategies managers are following when responding to dissatisfied customers plus a logistic regression analysis to examine what strategies are more effective in increasing subsequent rating of returning dissatisfied customers.

In the second step of Study–2, both positive and negative reviews and corresponding management responses are collected, processed, annotated, and classified. The third step of Study–2 involves the calculation of topic matching degree between and responses and reviews, construction of the model, and investigation of whether higher topic matching degree leads to greater rating growth on subsequent ratings of customers in a monthly timeframe.

In the last step of the research, the results are evaluated and discussed. After all, the research is finalized with the conclusion and future directions.

Figure 10 Research Design



3.6 Research Approach

The research approach is another fundamental component for both researchers and the philosophy of the research. Hence, this section describes deductive as well as inductive research approaches. Following, the triangulation of these approaches is introduced. Consequently, the adopted research approach is justified in the last part of this section.

3.6.1 The Deductive and Inductive Approaches

There are two common essential research approaches selected for research: deductive approach and inductive approach (Bryman & Bell, 2015). The deductive approach is usually linked with quantitative research by which the study is guided by theory (Collis & Hussey, 2013). Quantitative research examines objective theories to investigate the relationships amongst variables (Creswell & Creswell, 2018). The research commences with hypotheses and carries on with an empirical analysis to confirm or reject them in the deductive approach because it is connected with the positivist paradigm (Collis & Hussey, 2013). Ultimately, quantitative research uses experiments or surveys as data collection methods (Saunders et al., 2009).

Unlike the deductive approach, the inductive approach is generally linked with qualitative research, in which the study ends up with a theory (Collis & Hussey, 2013). The concept of qualitative research is concerned with the objective of exploring social or human problems through understanding individuals as well as groups. In contrast with the deductive approach, the study does not begin with hypotheses in this approach. Instead, research questions are used to narrow down the research scope and make inferences by findings; thus, it is linked with interpretivism paradigm (Bryman & Bell, 2015).

3.6.2 Abduction: Triangulation of Different Approaches

As Saunders et al. (2019) suggested, there is no strict line between deductive and inductive approaches. Abduction implies that it is probable to merge these approaches within the same research (Saunders et al., 2019). Some also defined that triangulating/mixing the quantitative and qualitative research methods, one can better understand the research problem (Creswell & Creswell, 2018). Especially in business and management contexts, using mixing approaches can yield remarkable outcomes as the potential weaknesses of one method could be made up for with the potential strengths of the other. Hence, abduction is common for business and management researchers (Saunders et al., 2019). The flexibility of the abductive approach makes it useful for researchers having a number of distinct research philosophies. It is argued that achieving pure deduction or induction might be quite difficult so numerous management researchers in practice use some abduction element. Nevertheless, the abductive approach is often associated with pragmatism and postmodernism (Saunders et al., 2019).

3.6.3 Adopted Research Approach

Different philosophies can lead researchers to different approaches, so positivists to deduction, interpretivism to induction, and pragmatists to abduction (Saunders et al., 2019). If the research topic has a rich source of literature that enables defining a theoretical framework and hypothesis, it tends to be more appropriate to adopt a deductive approach. However, if the research topic is new and has little source of literature, it might be more convenient to work inductively with data-driven knowledge. This implies generating and analysing data then revealing theoretical themes the data are suggesting. If the research topic has a rich source of literature in one context but far less in the other context, it is more proper to adopt an abductive approach in which data are used to find out a phenomenon, identify themes, and explain patterns to make it possible to create a new or modify an existing theory (Saunders et al., 2019).

Given the various assumptions explaining deductive, inductive, and abductive approaches, this research is conducted with an abductive approach. First, management responses have been widely examined in the literature. Second, as previously mentioned, the mixed method consisting of both quantitative and qualitative methods was adopted. The qualitative data was transformed into quantitative data after content analysis of responses and combined with other numerical measurements (response rate, response length, and response speed) for further statistical analysis (e.g., correlation analysis and the Welch's t-test). Thus, the adopted approach in this research is considered abduction.

3.7 Chapter Summary

This chapter adopts research philosophy and approach, outlined the research design, and introduced the deployed research methods. Various research philosophies were highlighted and why pragmatism was chosen for this research was explained. Next, the general research plan has been indicated in the form of a flow chart. Following, three different research approaches were introduced, along with the reason for choosing the abductive approach. After that, this chapter has introduced the strategy of this research. Last, the methods used in this research has been examined in detail.

CHAPTER 4:

STUDY–1: AN EXPLORATION OF STRATEGIC MANAGEMENT RESPONSES TO NEGATIVE REVIEWS

4.1 Management Responses to Negative Online Reviews

In online review sites, while numerous reviews are favourable, there exist a large proportion of unfavourable reviews that can remain on sites for a long term and bring along an enduring negative effect on business reputation (Hennig-Thurau et al., 2004; Lappas et al., 2016; Wang & Chaudhry, 2018) and performance such as room sales and online bookings (Min et al., 2015; Ye et al., 2009). Research has indicated that the damage caused by negative reviews is more than the benefit of positive reviews (Cui et al., 2012; Judith & Dina, 2006). One of the most extensively perceived challenges of businesses in searching for efficient ways to manage eWOM is to figure out if, when, and how to respond to online customer reviews, particularly negative ones (Sparks et al., 2016). Thus, as a form of public firm intervention provided by several platforms of social media, managerial response is a widespread way of coping with negative reviews in practice (Istanbulluoglu, 2017; Ma et al., 2015). Research suggested businesses respond to negative online reviews of customers to not be in a disadvantageous position and encouraged scholars to examine the effect of these responses (Leung et al., 2013; Min et al., 2015).

Studies have examined the content of the responses to negative reviews to detect and recommend the most optimum response strategies for organizations. Mostly, accommodative response versus defensive response strategy has been investigated. Based on Lee and Song (2010) 's statement, an accommodative response stands for a complete apology where a business admits the existence of the problem, takes the full accountability regarding the negative occurrence, states the regret, and attempts to relieve the damage. Conversely, a defensive response generally includes either justification where a firm repudiates the responsibility for the issue or an excuse in which a firm refuses the accused responsibility. Lee and Song (2010) point out that accommodative response strategy gives rise to better business evaluations when compared to defensive response or no response. Another study carried out by Lee and Cranage (2014) has shown that an accommodative response is more influential for high unanimity reviews while a defensive response is more influential for low unanimity reviews. More recently, Li et al. (2018) investigate semantically tailoring managerial responses to negative reviews' content and concentrated on the efficacy of accommodative ones versus defensive ones. Their research highlighted that an accommodative response to the product failure review (the review regarding the quality of the product) and a defensive response to the ordinary negative review (the review regarding only negative experiences) are efficacious at improving sales and customer purchase intents.

In the light of its immense value, the management response genre to negative reviews is widely investigated by scholars in business and linguistic fields in recent decades. Levy et al. (2013) analyse 255 managerial responses from 86 hotels in Washington, and they revealed eight moves: Active follow-up, Apology, Appreciation, Compensation, Correction, Explanation, Passive follow-up, and a Request for future patronage. They also find that high-rated hotels frequently respond to negative reviews with appreciation, apologies, and explanations. Sparks and Bradley (2017) categorize 150 responses belonging to highly rated and low rated hotels in Sydney, coming up with a typology called "Triple A Typology" consisting of Acknowledge, Account, and Action with subcategories per each. The Acknowledge category contains "Thank, Appreciate, Apologize, Recognize, Admit, Accept, and Dismiss;" the Account category includes "Excuse, Justify, Reframe, Penitential, and Denial;" the Action category consists of "Investigate, Referral, Rectify, Policy, Training, Direct Contact, and Compensate." They find that responses given by high rated often include the awareness of the firm about the occurrence of the event and appreciation for the customer's comment while responses given by low rated hotels often include denial of the event and the lack of rectification.

Zhang and Vásquez (2014) conduct a move structure analysis of 80 managerial responses to negative reviews posted by 4-star and 5-star hotels in China from the linguistic perspective. They identified 10 moves in responses: "Express Gratitude, Apologize for Sources of Problem, Invitation for a Second Visit, Opening Pleasantries, Proof of Action, Acknowledge Complaints/Feedback, Refer to Customer Reviews, Closing Pleasantries, Avoidance of Recurring Problems, and Solicit Response." Ho (2018) conducts a move structure analysis of 4256 responses from 2-star to 5-star hotels in different regions by approaching from the viewpoints of customers and defined the moves: "Acknowledging Problem, Continuing Relationship, Denying Problem, Expressing Feelings, Greeting, Recognizing Reviewer's/Comment's Value, Self-Promoting, and Thanking Reviewer." The research reveals that “Self- Promoting”, and “Denying Problems” are unpreferable, whereas the rest are preferable to customers.

More recently, Thumvichit et al. (2019) examine how high-ranked hotels deal with complaints on an anonymous platform to make a contribution to both the linguistic and travel industry. They identify six discrete moves, and their occurrence frequency in responses, which is shown in Table 2.

Table 2 Move Structure of Responses to Negative Reviews by Thumvichit et al. (2019*)*

Salutation	
Acknowledging Feedback	Giving gratitude - Valuing feedback - Stating apology or regret or concern
Brand Positioning	Specifying the commitment of the brand or Hotel’s norm
Coping with issues	Explanation of the reason caused to issue, Notifying the taken action, Acceptance of the Mistake
Concluding Remarks	Giving gratitude (2) - Stating apology or regret or concern (2) – Reinviting the customer – Demanding Direct Contact – Promising to enhance service
Closing	Details about the hotel and response provider

By using the analysis of Thumvichit et al. and other content analysis of responses to negative reviews, this study aims to understand how managers cope with negative reviews.

In the beginning, a comprehensive and common categorization of management responses was selected in light of the literature. The common categories include responses with a simple apology, compensation included responses, responses with the statement promising to improve service, brand positioning included responses, responses with the statement of admittance of mistakes, action included responses, explanatory responses, and responses including direct contact request. The distribution of these categories was manually analysed over a sample including 1000 responses. However, due to the lack of sufficient examples of responses with compensation statements and a multitude of responses with a simple apology, these categories were eliminated not to create imbalanced data set. Ultimately, the five response strategies were selected: action inclusive, the admittance of mistakes, brand positioning, explanatory, and direct contact requested. Considering these response strategies, this study explores response content strategies to negative online reviews; reveals correlation analysis between each response category and hotels' overall ratings, presents the differences of those strategies based on hotel star and overall rating. Second, this research points out what response strategies to negative reviews are more effective in increasing subsequent rating of dissatisfied returning customers.

4.2 Methods

4.2.1 Data Selection and Collection

Considering the particular significance of business-consumer interaction in the hospitality industry, this research uses a secondary data (TripAdvisor) from the hotel sector, which is publicly available on the internet. Data were collected on online customer reviews and management responses of 4-star and 5-star hotels in three popular tourist destinations in Europe -London, Paris, and Amsterdam- for the period from August 2010 to March 2018. Europe has become a major tourist destination, attracting 51% of worldwide visitor arrivals and 36% of international tourism revenue (Mitra et al., 2019). It is also the second-fastest growing regional destination after Asia and has the highest number of foreign visitors (Yasmeen, 2019). According to the report published by Euromonitor International, London was the most visited city and Paris was the second most visited city in Europe between

2016-2018. Amsterdam was the seventh most visited city in Europe between 2016 and 2018 (Geerts, 2018). The reason for choosing Amsterdam was to increase the diversity of data in addition to London and Paris, and to reduce the data extraction time to a shorter time due to the relatively a smaller number of 4-star and 5-star hotels. The reason for selecting 4-star and 5-star hotels on the data is to gain insight into the response strategies of high star hotels and make comparisons between them. Apart from the global popularity of these cities in the hospitality sector, another reason why these cities were chosen was the abundance of English reviews and responses they have on the social networking site from which the data was retrieved.

The data are collected with Python, which provides a wide range of efficient and useful libraries for data collection from websites. Two different data sets are created to store all the characteristics of reviews and responses and hotel characteristics separately. The relevant website allows customers to rate their experiences from 1 to 5. As this study focuses on the responses to negative reviews, the data are captured for the reviews with a rating of 1 to 3. A similar classification approach on customer ratings has been used in existing studies (Ku et al.; Park & Allen, 2013). First, all the negative reviews and responses to negative reviews of London, Amsterdam, and Paris 4-star and 5-star hotels were listed on the site when data collection were downloaded and stored in a data set. For customer reviews and management responses, all available information was collected, with the inclusion of user negative ratings, user negative reviews, review titles, user trip types, usernames (a username is unique and belongs to only one customer), review dates, responses to negative reviews, and response dates. Second, hotel characteristics were collected and stored in another data set, including hotel names, hotel starts, and cities where hotels located. After storing the data collected, a new data set was created by merging these two data sets.

The data set is large in volume and comprises both structured (e.g., review ratings, review and response dates, and hotel stars) and unstructured data (e.g., user reviews and management responses). All collected data are stored in both a comma-separated values (.csv) file and Apache CouchDB database in JSON format to use for data annotation software (see Data Annotation section). Apache Couch DB is an open-source documentoriented database using JSON to store data. Unlike a relational database, a CouchDB database does not use tables to store data and relationships. Instead, each database is an accumulation of independent documents.

Eventually, the raw data set involves 176,016 customer reviews and 111,189 attached managerial responses of 904 hotels located in London, Paris, and Amsterdam. Table 3 displays the total number of hotels, reviews, and responses in detail based on both locations and hotel stars.

Table 3 Collected Data

City	Total Number of Hotels	Total Number of Reviews	Total Number of Responses
	5-STAR		
London	126	22350	15324
Paris	90	6628	3933
Amsterdam	22	4181	3046
Total	238	33159	22303
	4-STAR		
London	341	91112	60213
Paris	461	33882	18840
Amsterdam	102	17863	9833

Total	904	142857	88886
Total	1142	176016	111189

4.2.2 Data Preparation

Data preparation is one of the significant steps within the data analysis process (Brownlee, 2020). Data preparation usually requires considerable effort and time since raw data generally comes with several imperfections such as inconsistencies, missing values, noise, and/or incompleteness. Because low-quality data leads to low-quality data analytics performance, data preparation has become a vital and principal stage in data science, machine learning, and the AI pipeline (Brownlee, 2020).

4.2.2.1 Integration of Data Sets

As mentioned in the Data Collection section, the data was collected for 4-star and 5-star hotels of London, Paris, and Amsterdam, creating six discrete data sets for customer reviews and responses. Initially, these six data sets were concatenated properly since all data sets have the same column header. Following this, the new data set was obtained, the data sets containing hotel information were merged, and the main data set was created.

4.2.2.2 Data Cleaning

Because this research focuses only on English responses and although the data was collected selecting English reviews and responses, it was important to control whether all reviews and responses are English. Therefore, using a language detection library in Python, languages of all the reviews and responses were checked. It was seen that 217 rows in the data set were not in English, and these rows were deleted from the data set.

In the next step, the data set was checked for missing values. It was detected that 8345 rows of the data did not contain trip type information of the customers. The trip type information of these rows was filled with a string as “unknown.” Besides, 38 rows of t data lacked username information of the customers. The usernames of those customers were manually found on the relevant website by refining the search criteria based on their review ratings and dates, and then inserted into data set.

4.2.2.3 Data Manipulation

A regular expression (or Regex) is used to avoid redundant strings in certain columns. Regex is a language construction that signifies a search pattern. Such patterns are usually used to find and replace the particular text (Lopez et al., 2014; Ramchand & Reiss, 2007). The column containing trip type information was converted into a categorical column removing redundant strings (e.g., “travelled with,” “travelled on”), and only trip type information (e.g., business, solo, family) was kept. The column comprising of response dates was including additional string with response date (e.g., “responded Jan. 20, 2017”), and just date information was kept. Last, the column including review ratings was refined from a combination of strings and special characters, and only review rating information was kept.

Data types of all columns in the data set were checked to make the data set suitable for analysis. First, the data types of review dates and response dates had been altered from string type to date type, and then a new column was created named response speed that refers to the difference between response date and review date.

4.2.3 Data Annotation

Theoretical and computational linguistics focuses on untangling the depths of language and getting hold of the computational properties of linguistic structures (Pustejovsky & Stubbs, 2012). Human language technologies (HLTs) aim to employ these insights and algorithms and transform them into functioning programs to enhance the interaction between us and computers (Pustejovsky & Stubbs, 2012). The vast amount of linguistic data available on the internet that are extraordinarily difficult to process by humans on their own has led linguistic modelling problems to be an important task in machine learning. Nonetheless, it is usually inadequate to provide a large computer the amount of data and assume it will learn linguistic patterns. As such, the data must be well prepared so that the computer can more simply detect patterns and inferences. This is mostly implemented by adding related metadata to a data set (Pustejovsky & Stubbs, 2012). During the annotation process, a metadata tag is used to assign elements of a dataset (Musen et al., 2015). Within text data annotation, data contain tags that signify criteria such as keywords, phrases, or sentences (Pustejovsky & Stubbs, 2012). The annotation should be accurate and related to the task, so machine/deep learning algorithms can train it efficiently. Poorly annotated data might lead a machine to weakly perform.

Prior to starting data annotation in this research, it is crucial to define how management responses given to negative reviews were going to be categorized pertinently not only to make a remarkable contribution to the relevant literature but also build an optimum training set for deep learning models, therefore, the subsections cover in which response categories are selected and how the sample data are annotated, respectively.

4.2.3.1 Defining Response Contents

Responding to customers' negative reviews may not singly be sufficient to abate the adverse impacts of those reviews (Li et al., 2018). In this sense, the content of management responses stands out as well as other characteristics of the responses. Research also demonstrates that the efficacy of management responses to negative reviews largely depends on the response's content (Min et al., 2015).

From the qualitative perspective, this research has been inspired by myriad studies that examined the content and structure of management responses to negative reviews (Levy et al., 2013; Li et al., 2018; Sparks & Bradley, 2017; Zhang & Vásquez, 2014). A combination of the response strategies that studies in the hospitality domain includes but is not limited to are apology, appreciation, compensation, correction, explanation, inviting the guest for a second visit, acknowledge of the failure, and a reference to action taken. As mentioned earlier, the structure of management responses was also examined from a linguistic perspective. Thumvichit et al. (2019) investigate how high-ranked hotels cope with customer complaints and define six discrete moves: acknowledging feedback, brand positioning, coping with issues, concluding remarks, and closing. They point out that coping with issues is the key component in handling complaints in terms of image protection and service recovery. They sectionalize coping with issues category into three subcategories as explaining causes of the incident, reporting action taken, and admitting mistakes.

In the light of the literature covering the content of managerial responses to customer complaints, a sample including randomly selected 1,000 responses was read and manually tagged according to these contents: apology, compensation, promising to improve service, brand positioning, the admittance of mistakes, action-included, explanatory, and direct contact requested. To avoid creating an imbalanced training set for the classification algorithm, the number of contents was reduced from eight to five contents. Because the contents of such apologies and promising to improve service were found in most of the responses, these contents are removed from data annotation. Compensation content was also removed from data annotation because in this content was found in very few responses. As a result of the reduction, this research defines five response contents: brand positioning, the admittance of mistakes, action inclusive, explanatory, and direct contact requested.

It is also worth noting that this study did not control the content of negative comments. It would be more efficient to analyse customers' dissatisfaction and determine which response strategies might be more appropriate to the relevant complaints.

4.2.3.1.1 Brand Positioning

Brand positioning commonly refers to building a proposition to get inside the mind of customers (Hooley et al., 2008). This comprises emphasizing the distinct features of a brand against its competitors and making them appealing to their customers (Koch & Gyrd-Jones, 2019). Not only are brand positioning strategies effective in establishing competitive advantage, but they are also beneficial in achieving organizational goals by developing and sustaining a positive image (Thumcivit et al., 2019). Such strategies are used in the tourism and hospitality industry as well to project a positive image of hotels. According to Thumcivit et al. (2019), brand positioning has a crucial role in a hotel’s reputation as it disseminates a favourable message about a hotel throughout the online community.

In this study, the responses are annotated according to the two steps that Thumcivit et al. (2019) identify for the brand positioning category: stating hotel's commitment and confirming hotel's standard. Stating hotel's commitment in a management response refers to an explicit statement implying what the hotel tries to achieve. Confirming the hotel's standard in a management response signifies how a manager states the hotel's standard despite the negative review with the aim of creating or sustaining a positive image. This commonly involves a statement that what customers experience does not reflect the real performance of the hotel. Some examples in the brand positioning category are shown below.

- "Our staff work hard to provide the best service to our guests, and it is always our main privilege to achieve this goal…"
- "We always strive for providing 5-star service at this hotel…"
- "It is clear from what you have written that your stay was not to the usual high standard of the…"
- "This is no way a reflection of how much we value our customers within the hotel…"

4.2.3.1.2 Admittance of Mistakes

In admittance of mistakes, a manager admits that a negative experience of customers occurred because of the hotel's fault. More specifically, the admittance of mistakes in a management response implies that the hotel is responsible for the negative situation. It is worth emphasizing that this content is different from expressing an apology devoid of a statement of taking responsibility for customers' negative experiences.

Some examples in the admittance of mistakes category in management responses are indicated below.

- “I am disappointed to hear the team were unapologetic and were unable to handle this delicate situation and for this I sincerely apologize…”
- “The condition in which you found your bedding is completely unacceptable and please accept my sincere apologies…”
- “There is no excuse at all for the service failures you describe…”
- “We have reimbursed immediately the extra charges and again, big apologize for this mistake…”

4.2.3.1.3 Action Inclusive

Action inclusive responses involve statements regarding what hotel management have done/will do to resolve the incident. The statements in action inclusive responses include that the issues are solved while the customer is at the hotel, or they will be solved to avoid having the same issues later. The actions in management responses consist of referring the issue to the relevant department of the hotel, changing some concrete aspects of a product or service, staff training and enhancement, implementing renovation/redecoration, and providing alternatives about the incident that the customer is dissatisfied with.

Some examples of inclusive action responses are shown below.

- “Following your review, we have carried out re-training within the team to ensure that all members of staff are fully equipped to provide the high standards of service…”
- “The front office staff will be updated on the area so that they are able to give correct directions to transport links, alternatively they should be able to research such things on the internet to answer guest queries…”
- “I would like to assure you that the hotel has recently undertaken a program to installed additional glazing to prevent this kind of complaint, and we hope to have this completed very soon...”
- “Your comments have been shared with the cleaning management team to take immediate action…”

4.2.3.1.4 Explanatory

Basically, in an explanatory response, a manager aims to explain what gave rise to the incident by providing information or proof to explicate. That is, the underlying reason for the incident is explained to the customer to make it understandable. An explanatory response can vary in content. As such, it can include a policy about the relevant topic, a detailed elucidation regarding a product or functioning of the service, or any reason as to why the customer's requests were not met.

Some examples of explanatory responses are shown below.

- "The reason why the rates were different with your friends booking is because this package does not have a set rate and is subject to availability."
- "We ask all our guests for a deposit; this useful information is being shown when making your booking at your Agency and is meant for incidentals such as booking excursions or a drink on your room."
- "Due to the grade two listing of the property a lift is not possible, this is something which we are transparent about over our website."
- "We are sorry for noise disturbance caused by fire alarm. However, according to the hotel policy, it is mandatory to conduct a fire drill at certain intervals."

4.2.3.1.5 Direct Contact Request Inclusive

In direct contact request inclusive responses, a manager would like to communicate with the customer individually via email or phone. However, not every response, including the direct contact statement, is tagged as direct contact with inclusive response. That is because, in some responses, managers express to customers that they can reach the hotel management if customers demand. Ultimately, only the responses that indicate that managers who request direct contact with customers are annotated as direct contact request inclusive responses.

Some examples of direct contact request inclusive responses are shown below.

- "I am aware of the problems you faced. Can you please contact me via my email? As I can look into the matter personally."
- "I would appreciate the opportunity to discuss this directly and put this right for you. Please contact me via…"
- "Dear customer, I am really glad that we had a phone call with you right after your check out, and we sorted the problem out."

4.2.3.2 Annotating Data

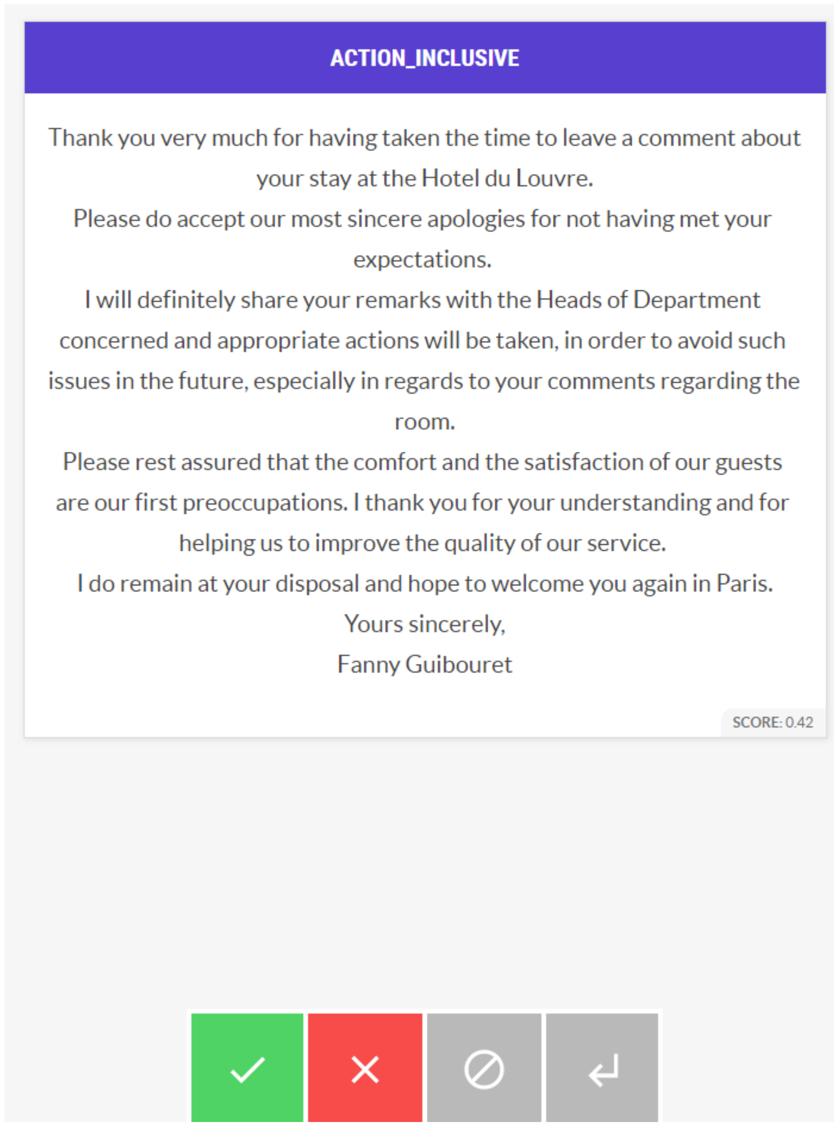
In this research, a web-based annotation tool, Prodigy, is used to annotate data. Prodigy provides multiclass, multilabel, and binary annotation options. Although the nature of the text classification task in this research seems a multilabel classification task (since one response can include multiple categories), the binary classification approach for each response content was employed. Furthermore, to reduce potential error and avoid confusion among five categories during the annotation process, we prefer to concentrate on only one content at one time. As such, they are providing only a single label, which makes the annotation decision much simpler by just accepting or rejecting an example based on the category.

Another advantage of binary annotation in Prodigy is that it provides active learning and a model in the loop. When we start the server for annotation, the application brings binary questions, and as we annotate, the model updates itself with our answers and suggests its predictions with a score. In addition, the tool allows us to provide match patterns that contain words or phrases related to the given label, for instance, "mistaken" or "unacceptable" in the admittance of mistakes category, or "strive for" and "high standards" in brand positioning category. The pattern matches are also getting integrated with the model suggestions, enabling the model to begin with enough examples to make better suggestions.

Figure 11 shows an example of text to annotate. The buttons at the bottom of the image gives the options about what to do with the text. The green button refers to accepting this training example, so if we think that the example involves the category for the annotation, we click this button. The red button refers to rejecting this example so that we click this button if we think the example text does not involve the category for the annotation. The grey button located next to the red button refers to ignoring this example; we click this button in case the example text is irrelevant to our task, or it is confusing to accept or reject. Last, the grey button on the far right is the undone button that is used when we want to revise or change the previously annotated text.

A set of randomly selected responses are uploaded into Prodigy, and the responses are labelled separately for each defined response category. Thus, five different data sets are created; each data set contains a thousand responses, 500 of which are accepted and 500 of which are rejected, according to the relevant category. All these data sets have been stored in a database for later use as training sets in text classification.

Figure 11 Screenshot of Prodigy Annotation Tool



4.2.4 The Adoption of Text Categorizer

In this research, spaCy's TextCategorizer, CNN, logistic regression, and SVC were employed for automatic text classification. As a free and open-source library for advanced natural language processing in Python, spaCy allows users to develop applications that process and grasp large volumes of text, as described by its creators. Several efficient features make spaCy one of the world leading libraries for NLP: linguistically motivated tokenization, NER, POS, text classification, pretrained word vectors, simple integration with custom deep learning models such as PyTorch, TensorFlow and other frameworks, industrial-level speed, support for more than 60 languages, and easy model packaging, and deployment and workflow management.

TextCategorizer is spaCy's text classifier, a trainable pipeline component using supervised learning to predict categories over a whole document. The training phase requires examples with their class labels, namely, a labelled data set. TextCategorizer is available for both single-label and multilabel classifiers. Because the data set in this research is annotated for each response content separately, the single-label classifier model of TextCategorizer was employed for each data set created after data annotation. TextCategorizer enables developers user-friendly and end-to-end approaches, so users do not directly have to cope with the neural network architecture. However, we can simply and optionally change the model architecture. In this research, the default neural network architecture of spaCy's TextCategorizer, CNN, Logistic Regression, and SVC methods were adopted and performed on the same data sets to automatically classify responses based on the related categories. Methods producing the best accuracy were selected as classifiers for the corresponding categories.

4.3 Exploratory Data Analysis

The data consists of 176016 rows where 120062 negative reviews are responded to, whereas 55,954 negative reviews are not responded to, whereas the total number of 4-star and 5-star hotels located in Amsterdam, London, and Paris is 1,142.

As Figure 12 indicates, the number of 5-star hotels is the greatest in London with 126, whereas it is 90 and 22 for Paris and Amsterdam. However, the number of 4-star hotels is the greatest in Paris with 461, whereas it is 341 and 102 for London and Amsterdam, in turn.

Figure 12 Number of Hotels in Cities

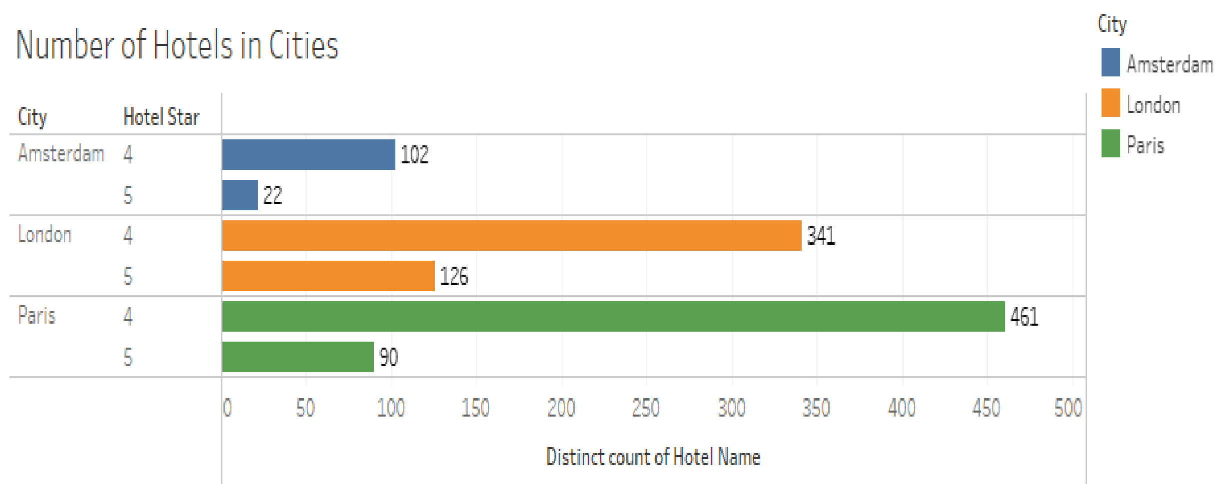


Figure 13 below shows the percentile of negative reviews in total reviews considering different cities and hotel stars, the greater circle, the greater percentile of negative ratings. Regarding 5-star hotels, 17.28% of total reviews are negative in London hotels, while 16.52% and 16.34% of total reviews are negative for Amsterdam and Paris. Looking at the 4-star hotels, Paris hotels have the greatest percentile of negative reviews with 25.77%, where 23.27% and 18.63% of the total reviews are negative for London and Amsterdam, respectively.

Figure 13 Percent of Negative Reviews

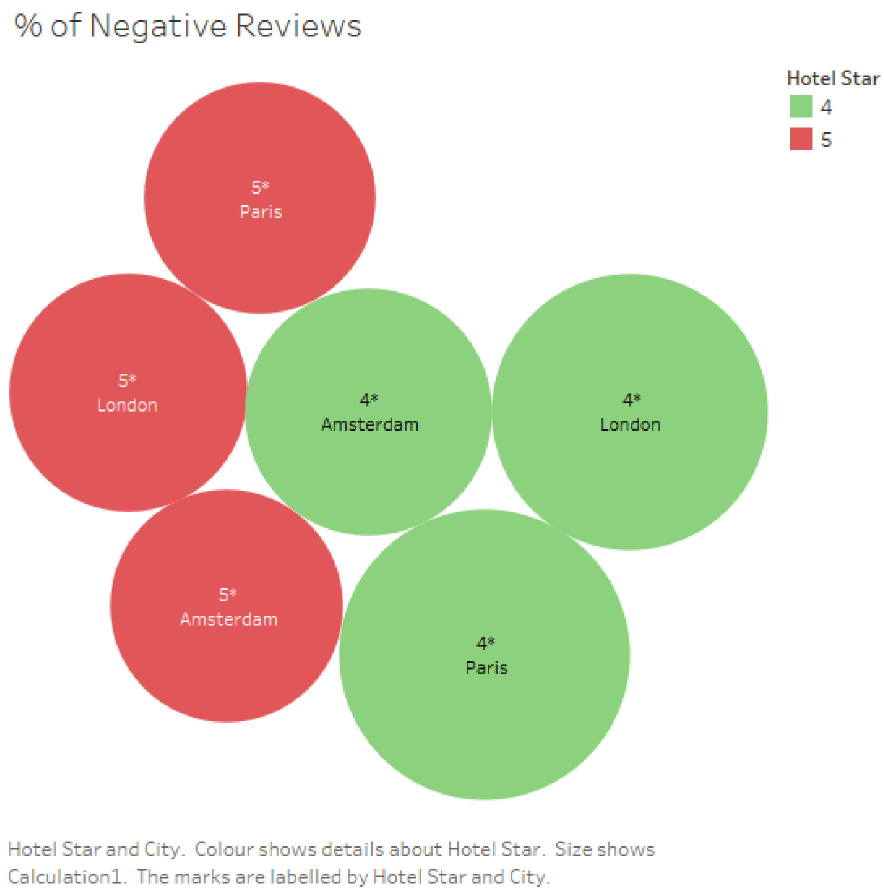
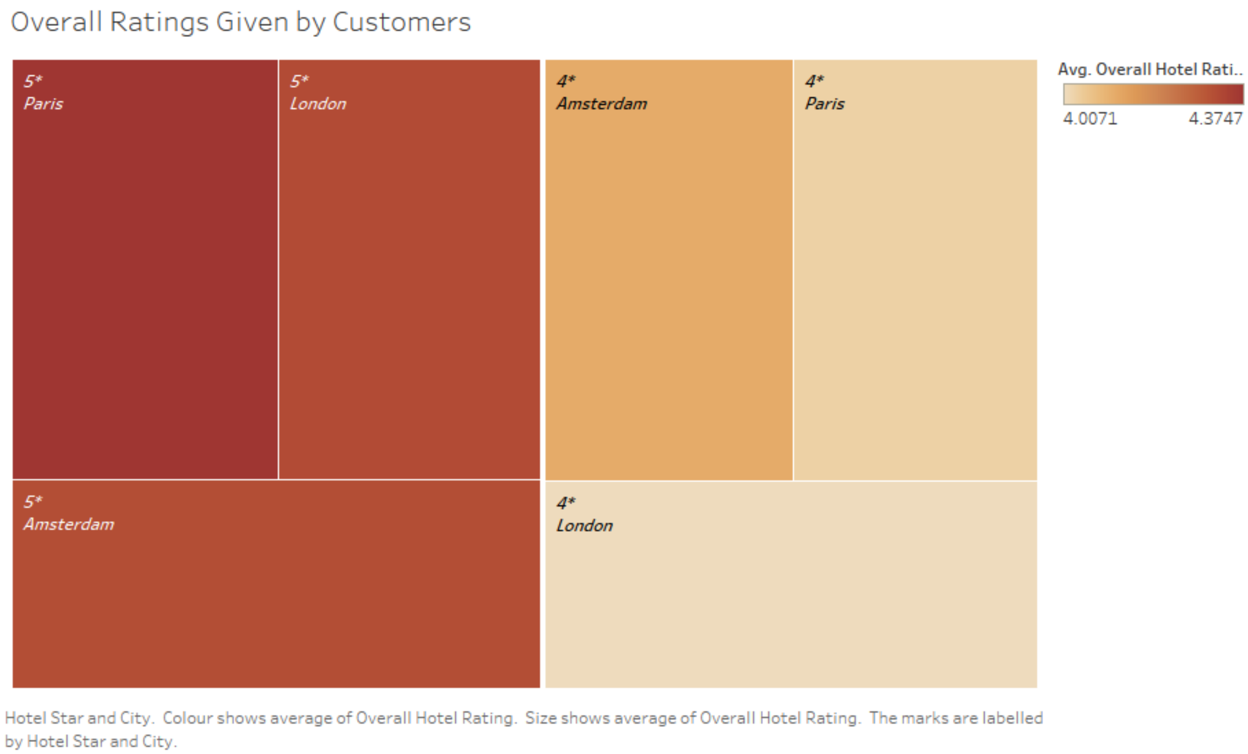


Figure 14 shows the heatmap of overall ratings given by customers, comparing cities and hotel stars. Overall ratings of hotels are calculated considering the aggregation of multiplied ratings based on the rating type and the number of reviews. The calculation is made with the formula:

$$\frac{5 * Excellent + 4 * Very\ Good + 3 * Average + 2 * Poor + 1 * Terrible}{(Excellent + Very\ Good + Average + Poor + Terrible)}$$

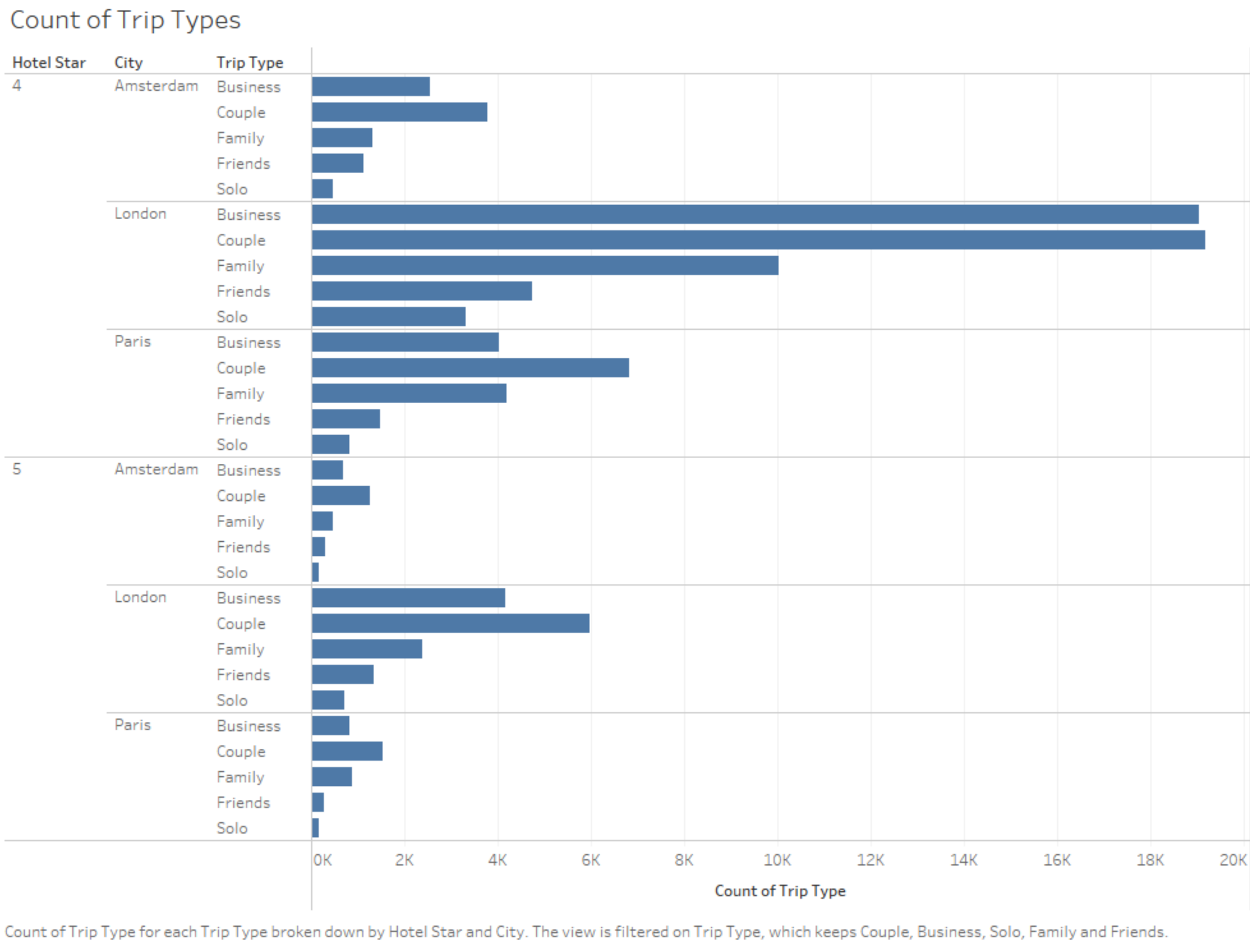
Regarding 5-star hotels, the highest average of overall ratings belongs to the hotels located in Paris with 4.3747, in which hotels in London and Amsterdam have 4.3237 and 4.3185 average overall ratings, respectively. In terms of 4-star hotels, the highest average of overall ratings is 4.1121 for hotels in Amsterdam, where hotels in Paris and London have 4.0307 and 4.0071 average overall ratings.

Figure 14 Comparison of Overall Hotel Ratings



Examining the count of customers' trip types (see Figure 15), it is highly noticeable that the distributions of the count of trip types are quite similar for both 4-star and 5-star hotels and all cities. For each city and hotel star, the number of customers visiting the hotels as couples is the highest; the number of customers visiting the hotels for business purposes is the second highest; the number of customers staying at the hotels with their family is the third; the number of customers staying at the hotels with their friends is the forth; and lastly, the number of single customers staying at the hotels in the lowest when compared to other customers having different trip types.

Figure 15 Distribution of Trip Types



4.3.1 Response Characteristics

4.3.1.1 Response Rate of Hotels

When looking at the response rate of hotels based on their stars (see Figure 16), the response rate of 5-star hotels is almost 6% higher than the response rate of 4-star hotels, considering the aggregate number of hotels in the given cities.

Figure 16 Response Rates Based on Hotel Stars

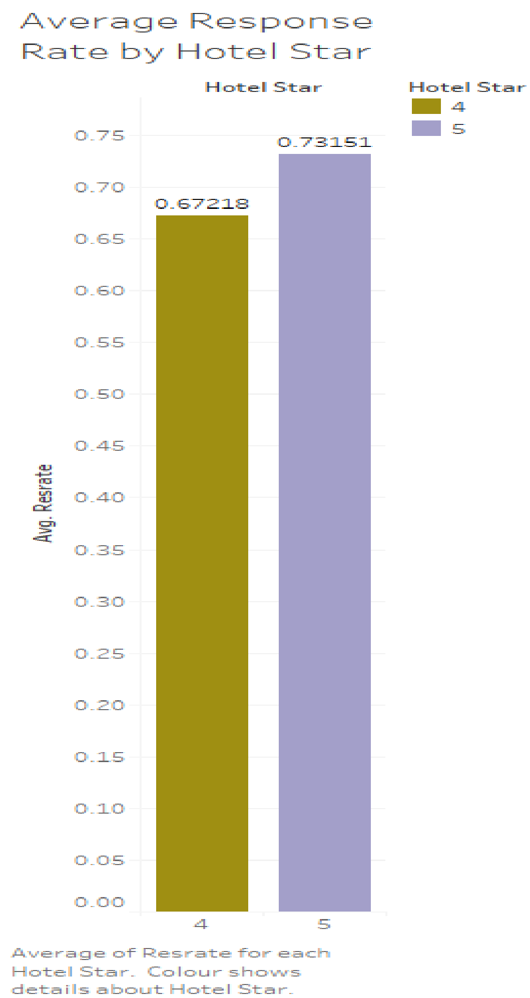
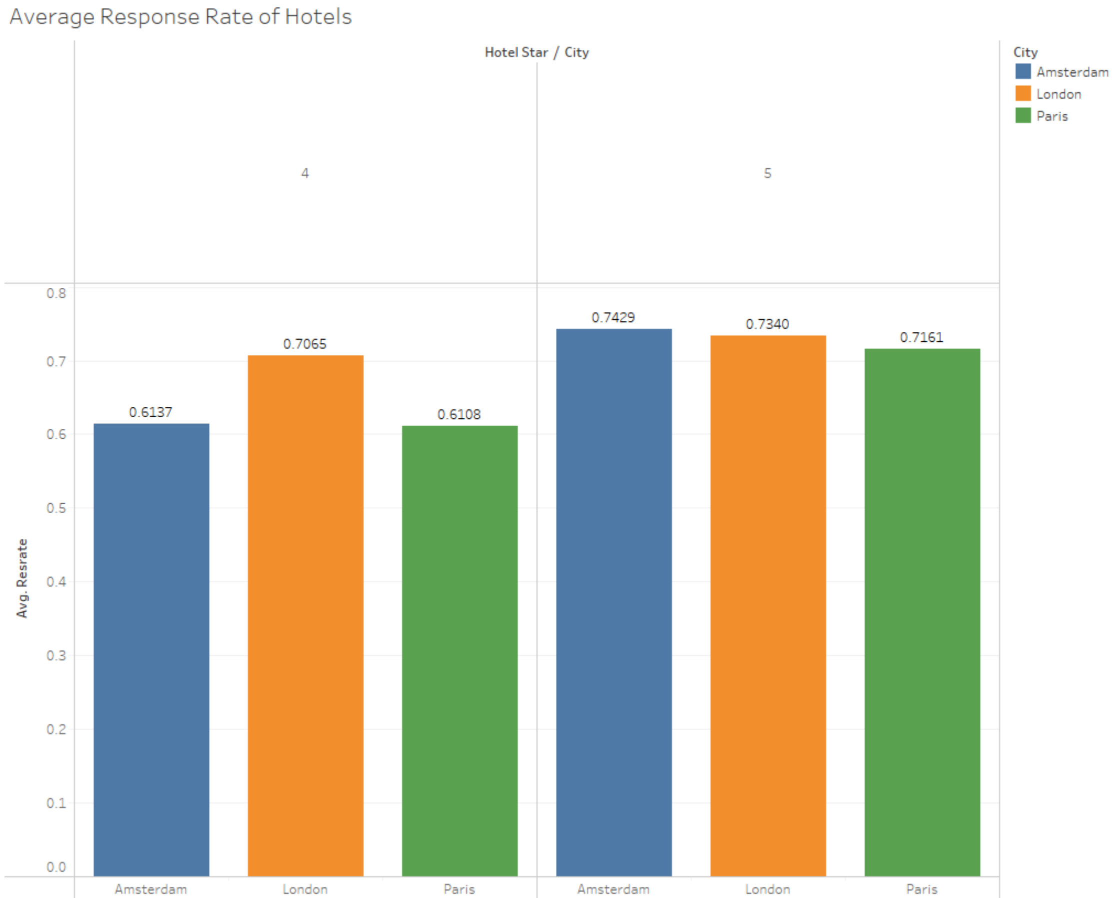


Figure 17 shows the average response rate of hotels based on both cities and hotel stars. As for the 5-star hotels, average response rates of hotels are quite close to each other with almost 74%, 73%, and 71% for Amsterdam, London, and Paris, respectively. However, there is a substantial difference between average response rates of hotels for 4-star hotels. While the average response rates are about 61% for Amsterdam and Paris, it is nearly 71% for hotels located in London.

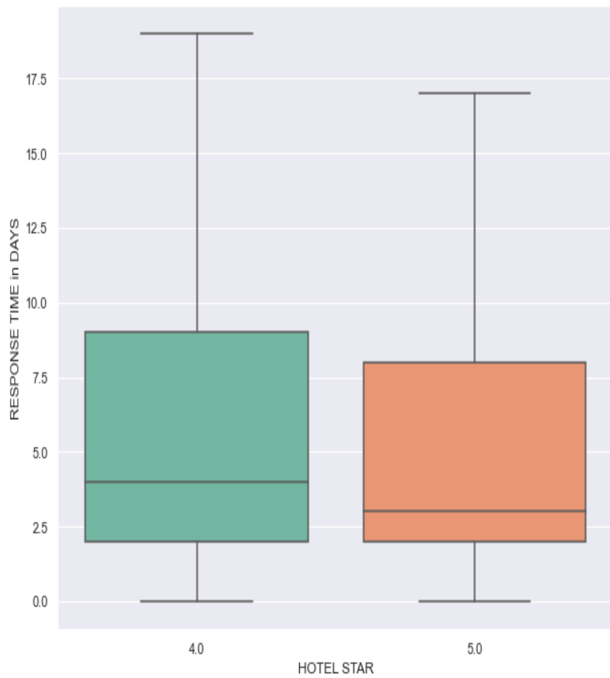
Figure 17 Average Response Rate Based on Cities and Hotel Stars



4.3.1.2 Response Time in Days (Response Speed)

Figure 18 indicates the comparison of response time distribution in days based on hotel stars with outliers being removed to avoid a noisy visualization in the boxplots. Because some responses were given after years, it significantly increases the average response times as 30.64 and 42.82, and standard deviations as 226.50 and 327.0 for 4-star and 5-star hotels, respectively. That is because, it might be more appropriate to compare the medians of response times. Looking at the boxplots and table, the median of the response time of 4-star hotels is 4.0 days whereas it is 3.0 days for 5-star hotels. As such, the median of the response time of 5-star hotels is 1.0 day less than the median of response time of 4-star hotels. The minimum response time for both hotel types is 0.0 while maximum response times are 5595.0 and 5133.0 for 4-star and 5-star hotels in turn. In terms of distribution of response time of 4-star hotels, 25% are between 0 and 2.0 days, 50% are between 0 and 4.0 days, 75% are between 0 and 9.0 days, and the rest are above 9.0 days. The distribution for 5-star hotels indicates that 25% are between 0 and 2.0 days, 50% are between 0 and 3.0 days, 75% are between 0 and 8.0 days, and the rest are above 8.0 days.

Figure 18 Response Speed Comparison of 4-star and 5-star Hotels



Hotels	4-star	5-star
count	88886	22303
Response Time Mean	30.64	42.82
Response Time St.Dev	226.50	327.07
Response Time Median	4.0	3.0
25%	2.0	2.0
50%	4.0	3.0
75%	9.0	8.0
min	0.0	0.0
max	5595.0	5133.0

Figure 19 shows the medians of response times in days based on the cities and hotel stars. Regarding 4-star hotels, the medians of the response time of the hotels in London and Amsterdam are the same at 3 days, whereas it is 5 days for the hotels located in Paris. Looking at the 5-star hotels, the medians are the same at 3 days for the hotels in Amsterdam and London while it is 4 days for Paris hotels.

Figure 19 Medians of Response Speed

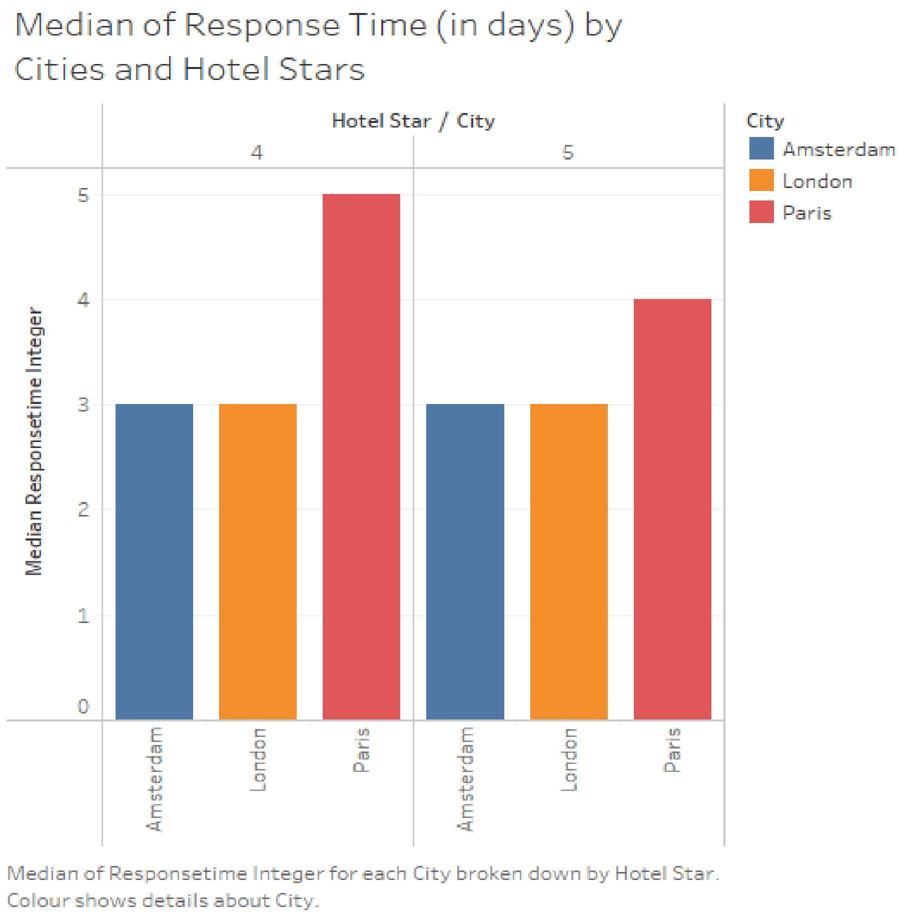
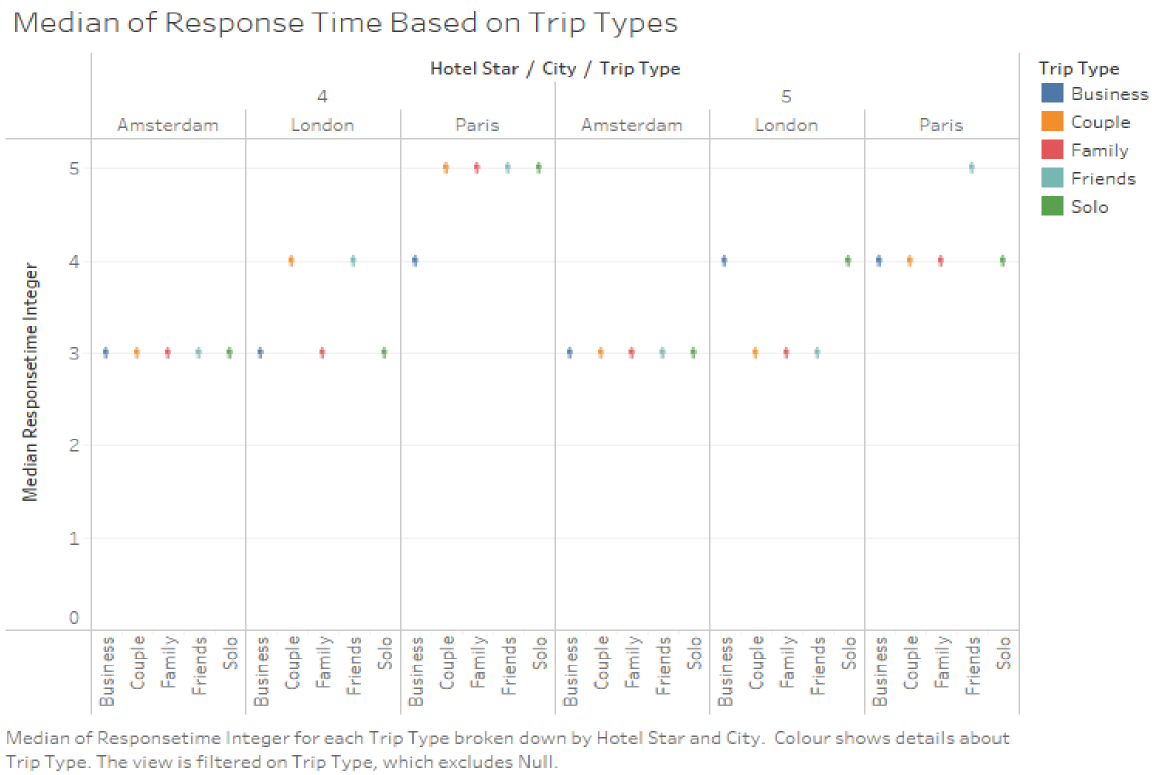


Figure 20 represents the medians of response times in days concerning customers' trip types which are travelling for business reasons, travelling with the couple, family, or friends, and travelling solo. Regarding Amsterdam, the medians for all trip types are 3 days for both 4-star and 5-star hotels. Considering London 4-star hotels, the median is 3 days for trip types of business, family, and solo while it is 4 days for trip types of couple and friends. Looking at the London 5-star hotels, the median is 3 days for trip types of couples, family, and friends while it is 4 days for trip types of business and solo. When Paris 4-star hotels are examined, the median is 4 days for the trip type of business in which it is five days for the rest of trip types. Regarding 5-star hotels in Paris, it is 5 days for the trip type of friends where the medians are 4 days for the rest of trip types.

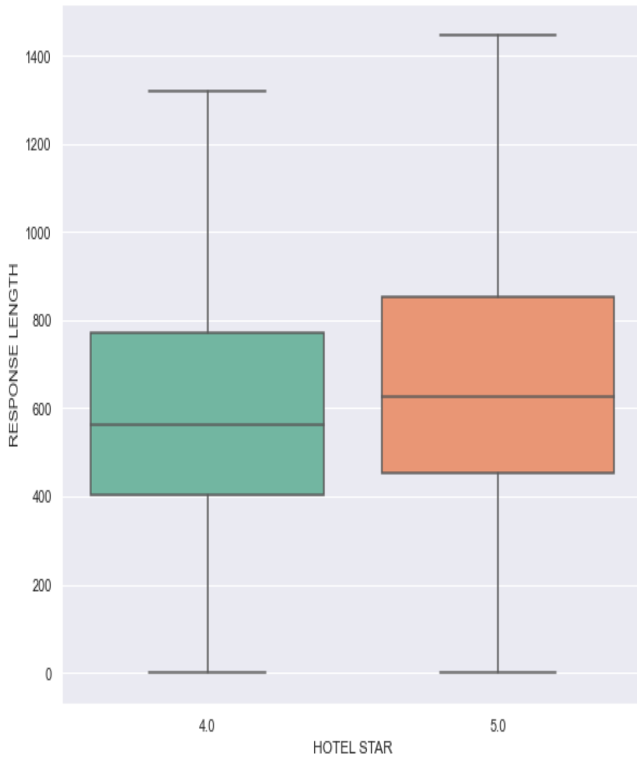
Figure 20 Medians of Response Speed Based on Trip Types



4.3.1.3 Response Lengths

Figure 21 indicates the comparison of the distribution of response lengths (the count of characters in responses) based on hotel stars, with outliers being removed to avoid a noisy visualization in the boxplots. The means of response lengths are 630.62 and 706.99 where the standard deviations are 348.43 and 405.99 for 4-star and 5-star hotels, respectively. Thus, the medians of response lengths are 562 for 4-star hotels and 627 for 5-star hotels. Regarding 4-star hotels, 25% of responses include up to 405 characters, 50% contain slightly 400 to 800 characters, and the rest of 25% consist of more than 750 characters. As for 5-star hotels, 25% of responses include up to 454 characters, 50% contain almost 450 to 850 characters, and the rest 25% consist of more than 850 characters.

Figure 21 Response Length Comparison of 4-star and 5-star Hotels



Hotels	4-star	5-star
count	88886	22303
Response Length Mean	630.62	706.99
Response Length Standard Dev.	348.43	405.99
Response Length Median	562	627
25%	405.00	454.00
50%	562.00	627.00
75%	771.00	851.00
max	8673.00	7337.00

Comparing three cities in Figure 22, average response lengths provided by 4-star hotels are close to each other with 633.5, 626.1, and 643.6 for Amsterdam, London, and Paris, respectively. While the average response lengths provided by 5-star hotels in Amsterdam and Paris are quite close to each other with slightly over 750 characters, the average response length provided by managers in 5-star London hotels is slightly over 680 characters. Comparing the medians of response lengths, there is a negligible difference for 4-star hotels with 557, 560, and 569 for Amsterdam, London, and Paris, in turn. However, the median values are quite close to each other for 5-star hotels located in Amsterdam and Paris with 673 and 683 while it is 600 for 5-star London hotels.

Figure 22 Response Length Comparison Based on Cities

4.3.1.4 Response Categories

All the response categories were classified using four different models including two traditional machine learning models that are support vector machines and logistic regression, two deep learning models that are spaCy and CNN.

For all models, annotated data sets for each category were split into training sets and test sets as 70% and 30%, respectively. Because each data set contains 1,000 annotated samples, classifications are made based on training sets with 700 samples, and test sets with 300 samples. Accordingly, the models with the highest classification accuracy and F-score are selected for classification for each category separately. That is, for each response topic, the performances of all the classification approaches abovementioned are tested in order. As a result, the classifier with the best accuracy for each response category is used to classify all response data set for the relevant category.

Table 4 shows the classification comparison for action inclusive responses. For this category, spaCy's TextCategorizer outperforms the other models with 0.81 accuracy and 0.84 F-Score. Therefore, the model created with Spacy is used on the large data set to classify action inclusive responses.

Table 4 Classification Comparison for Action Inclusive Responses

Model	Accuracy	Precision	Recall	F-Score
SVC	0.73	0.74	0.73	0.74
Logistic Regression	0.74	0.75	0.74	0.74
Spacy	0.81	0.79	0.88	0.84
CNN	0.79	0.76	0.77	0.76
Best	**0.81**			**0.84**

Table 5 indicates the classification comparison for admittance inclusive responses. For this category, the best accuracy that is 0.83 and F-score that is 0.83 were obtained with the CNN model. Therefore, the model created with CNN was used on the large data set to define admittance inclusive responses.

Table 5 Classification Comparison for Admittance Inclusive Responses

Model	Accuracy	Precision	Recall	F-Score
SVC	0.70	0.71	0.70	0.70
Logistic Regression	0.69	0.71	0.69	0.69
Spacy	0.72	0.70	0.81	0.75
CNN	0.83	0.83	0.83	0.83
Best	**0.83**			**0.83**

Table 6 indicates the classification comparison for brand positioning inclusive responses. For this category, CNN performed better than other models with 0.88 accuracy and 0.87 F-Score Therefore, the model created with CNN was used on the large data set to define brand positioning inclusive responses.

Table 6 Classification Comparison for Brand Positioning Inclusive Responses

Model	Accuracy	Precision	Recall	F-Score
SVC	0.78	0.78	0.78	0.78
Logistic Regression	0.77	0.77	0.77	0.77
Spacy	0.86	0.80	0.92	0.85
CNN	0.88	0.87	0.88	0.87
Best	**0.88**			**0.87**

Table 7 indicates the classification comparison for direct contact request inclusive responses. For this category, the best accuracy that is 0.88 and F-score that is 0.89 were obtained with the CNN model. Therefore, the model created with CNN was used on the large data set to define direct contact request inclusive responses.

Table 7 Classification Comparison for Direct Contact Request Inclusive Responses

Model	Accuracy	Precision	Recall	F-Score
SVC	0.85	0.86	0.85	0.85
Logistic Regression	0.80	0.83	0.80	0.80
Spacy	0.88	0.86	0.90	0.88
CNN	0.88	0.87	0.90	0.89
Best	**0.88**			**0.89**

Table 8 shows the classification comparison for explanatory responses. For this category, spaCy's TextCategorizer outperforms the other models with 0.85 accuracy and 0.88 F-Score. Therefore, the model created with Spacy is used on the large data set to define explanatory responses.

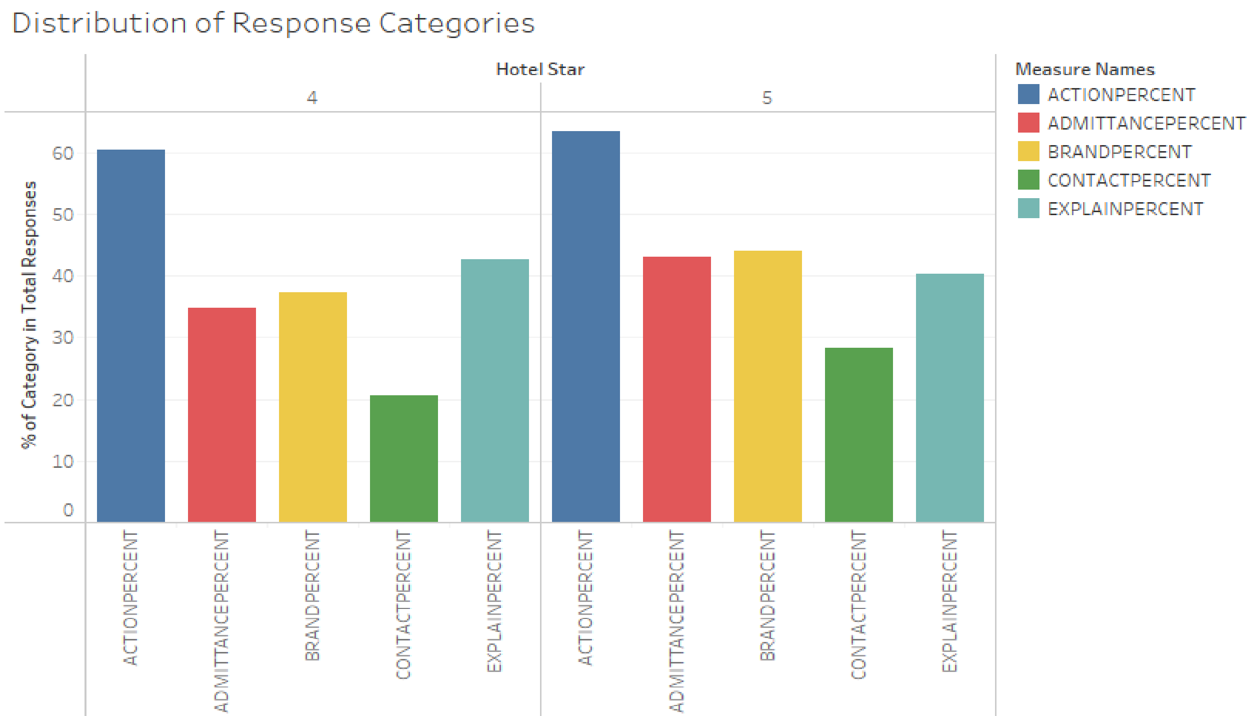
Table 8 Classification Comparison for Explanatory Responses

Model	Accuracy	Precision	Recall	F-Score
SVC	0.79	0.79	0.79	0.79
Logistic Regression	0.77	0.78	0.77	0.77
Spacy	0.85	0.83	0.92	0.88
CNN	0.83	0.83	0.86	0.84
Best	**0.85**			**0.88**

4.3.1.4.1 Distribution of Response Categories

Figure 23 indicates the distribution of response categories in percent for 4-star and 5-star hotels for all cities. Regarding 4-star hotels, slightly over 60% of responses include Action category. 42.5% of responses contain, Explanatory category, 37.2% of responses include Brand Positioning category, 34.7% of responses consist of Admittance category, and 20.5% of responses include Direct Contact Request category. Looking at the 5-star hotels, almost 64% of responses contain Action category. Unlike the response category distribution in 4-star hotels, Brand Positioning category is the second most common category in the responses and provided 5-star hotels with about 44%. Following, about 43% of responses include Admittance category while 40.2% of responses involve the Explanatory category. Last, 28.3% of responses contain Direct Contact Request category.

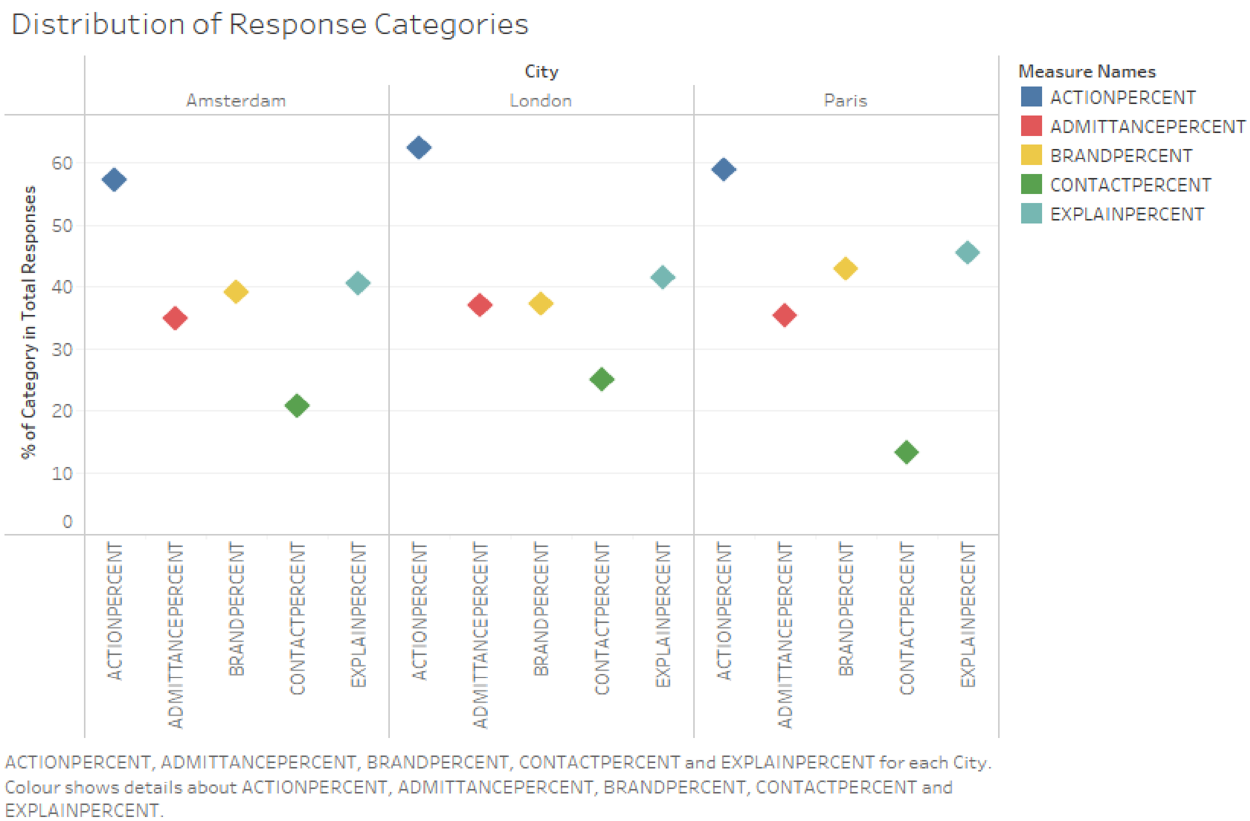
Figure 23 Distribution of Response Categories Based on Hotel Stars



ACTIONPERCENT, ADMITTANCEPERCENT, BRANDPERCENT, CONTACTPERCENT and EXPLAINPERCENT for each Hotel Star. Colour shows details about ACTIONPERCENT, ADMITTANCEPERCENT, BRANDPERCENT, CONTACTPERCENT and EXPLAINPERCENT. The view is filtered on Hotel Star, which keeps 4 and 5.

When distribution of response categories in percent based on different cities were examined (see Figure 24), we are seeing a similar pattern in terms of sequence of categories. As such, the most seen category in responses in all cities is Action where Explanatory, Brand Positioning, Admittance, and Direct Contact Requested categories come in a descending order. The Action category is seen with the highest percent in London hotels with almost 62.3% while it is 58.8% and 57.2% for Paris and Amsterdam, respectively. The Explanatory and Brand Positioning categories exist in responses provided by Paris hotels with the highest percent are 45.3% and 42.8%, respectively, where they are 41.3% and 37.2% for London and 40.4%, and 34.8% for Amsterdam hotels. The Admittance category is seen with the highest percent in London hotels with 37% while it is 35.3% for Paris hotels and 34.8% for Amsterdam hotels. Lastly, the Brand Positioning category is seen with the lowest percent in Paris hotels with 13.2%, whereas it is 20.7% for Amsterdam hotels and 25% for London hotels.

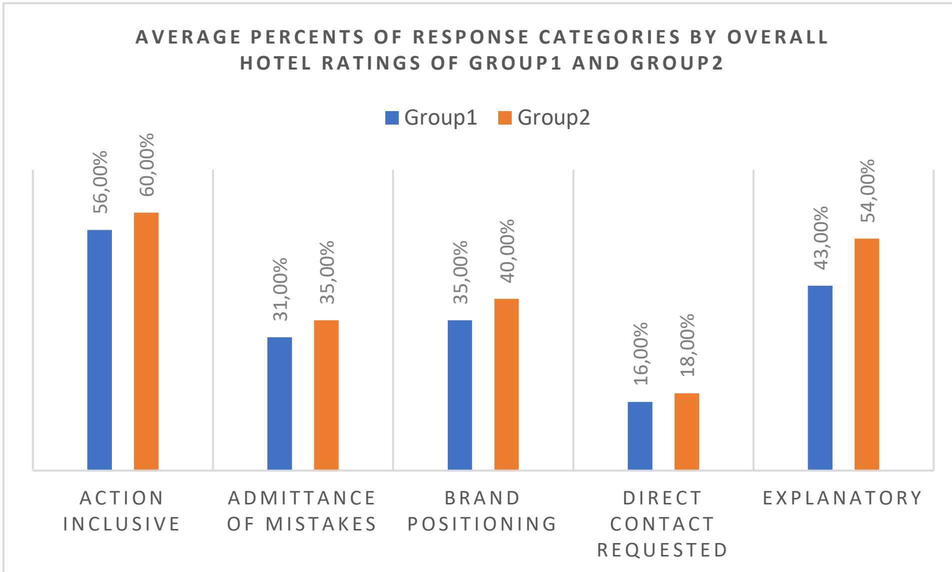
Figure 24 The Distribution of Response Categories Based on Cities



4.3.1.4.2 Comparison of Percent of Response Categories Based on Overall Hotel Ratings

To investigate the comparison of percentage of categories in responses, the quartile approach for overall hotel ratings was adopted. A hotel's overall rating belongs to Group1 if it is in Q1 and Q2, otherwise it belongs to Group2 (Q3 and Q4). More specifically, Q1 and Q2 are the quartiles that below the median of overall ratings, while Q3 and Q4 are the quartiles that above median of overall ratings. Total number of hotels in Group1 is 275 while it is 666 in Group2. Figure 25 shows the percent of each response category for Group1 and Group2.

Figure 25 The Comparison of Response Categories Based on Overall Ratings



The average percent of each category in responses in Group2 is greater. The average of action included responses are 56% and 60%, the admittance of mistake(s) included responses are 31% and 35%, brand positioning included responses are 35% and 40%, direct contact requested responses are 16% and 18%, and explanatory responses are 43% and 54% for Group1 and Group2, respectively.

Then the differences were tested with Welch's t-test. Table 9 shows important statistical differences between Group2 and Group1 for the categories of action included, the admittance of mistake(s), brand positioning, and explanatory where the confidence interval is 95% and p values < 0.05. However, there is no significant difference for the category of direct contact requested as $p > 0.05$

Table 9 Welch's t-test for the Comparison of Group2 and Group1

Categories	T	p-value	Confidence Interval (95%)

Action Inclusive	2.882743	0.00	[0.01, 0.06]
Admittance of Mistake	3.975366	0.00	[0.02, 0.07]
Brand Positioning	3.551795	0.00	[0.02, 0.07]
Direct Contact Requested	1.855877	0.06	[-0.0, 0.04]
Explanatory	6.861705	0.00	[0.08, 0.14]

4.3.1.4.3 Correlations of Response Categories With Overall Hotel Ratings

Table 10 shows the correlations between each response category and the overall hotel ratings. Looking at the p values, the categories action inclusive, admittance of mistakes inclusive, and direct contact requested responses have significant relationships (where p=0.000 for each) with the overall hotel ratings. However, there is no significant relationship for action inclusive and explanatory responses with overall hotel rating. The correlation coefficient *r* is the highest with 0.264 for the correlation of direct contact requested responses and the overall hotel rating while it is 0.21 for the correlation of admittance of mistake inclusive responses and the overall hotel rating and 0.195 for the correlation of brand positioning inclusive responses and the overall hotel rating.

Given that it might be said that there is a weak positive relationship between brand positioning inclusive responses and the overall hotel rating, there is a weak positive relationship between admittance of mistakes inclusive responses and overall hotel rating, and there is a weak positive relationship between direct contact requested responses and overall hotel rating. However, there is no relationship for action inclusive and explanatory responses with overall hotel rating.

Table 10 Correlations of Response Categories and Overall Hotel Rating

Independent Variable	Dependent Variable	Method	Correlation Coefficient(r)	p-value
Action Inclusive	OverallHotelRating	Pearson	0.055	0.252

Admittance of Mistake	OverallHotelRating	Pearson	0.210	0.000
Brand Positioning	OverallHotelRating	Pearson	0.195	0.000
Direct Contact Requested	OverallHotelRating	Pearson	0.264	0.000
Explanatory	OverallHotelRating	Pearson	0.087	0.087

4.3.1.5 Correlations of Response Characteristics with Overall Hotel Ratings

Table 11 indicates the correlation table between different response characteristics and overall hotel ratings, which is also shown as a correlation matrix on Figure 26. The response characteristics refer to independent variables on the table as response rate, response speed (in days), and response length while the overall hotel rating is the dependent variable. The correlation coefficient *r* is almost 0.27 for the correlation of ResponseRate and OverallHotelRating where p value (*p-corr)* is 2.63661e–16, which is far less than 0.05, and 95% confidence interval (*CI95%*) is 0.21, 0.33. The correlation coefficient *r* is almost 0.05 for the correlation of ResponseSpeed and OverallHotelRating where p value (*p-corr)* is 8.100527e-01, which is far greater than 0.05, and 95% confidence interval (*CI95%*) is -0.01, 0.11. The correlation coefficient *r* is above 0.30 for the correlation of ResponseLength and OverallHotelRating where p value (*p-corr)* is 1.690967e–20, which is far less than 0.05, and the 95% confidence interval (*CI95%*) is 0.21, 0.33.

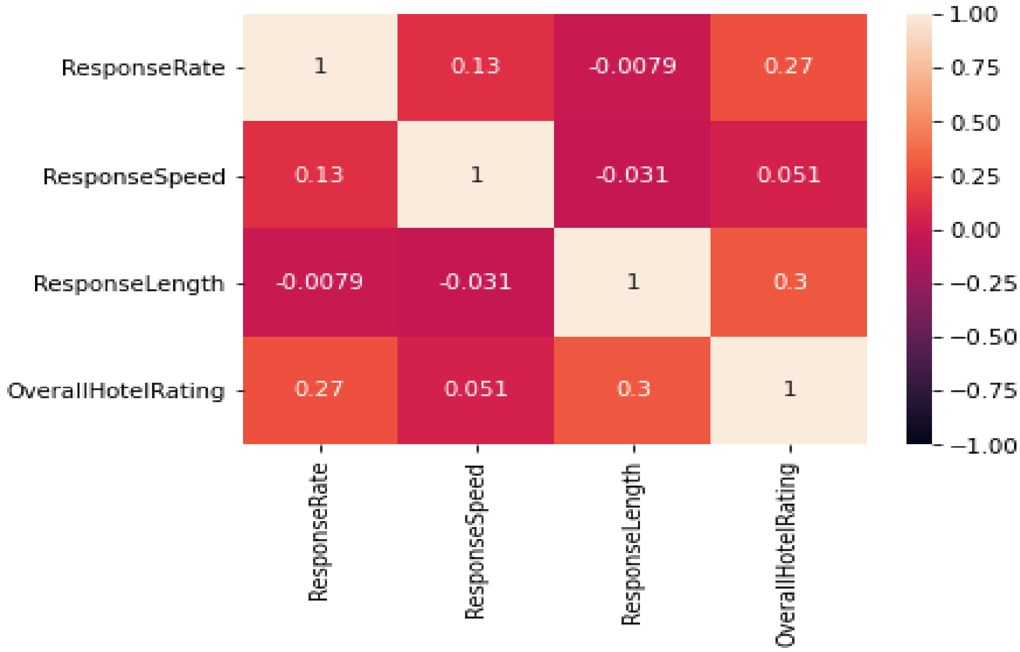
Given that it might be said that there is a weak positive relationship between response rate and overall hotel rating, there is a weak positive relationship between response length and overall hotel rating, and there is no relationship between response speed and overall hotel rating.

Table 11 Correlation Table

Independent Variable	Dependent Variable	Method	Correlation Coefficient(r)	Confidence Interval(95%)	p-value
ResponseRate	OverallHotelRating	Pearson	0.271	[0.21, 0.33]	0.00

ResponseSpeed	OverallHotelRating	Pearson	0.051	[-0.01, 0.11]	0.81
ResponseLength	OverallHotelRating	Pearson	0.304	[0.25, 0.36]	0.00

Figure 26 Correlation Matrix



4.4 Investigation of Responses to Repeated Customers

In the second phase of Study-1, responses provided to the repeated customers, who visit the same hotel after leaving negative rate, are investigated to be able to figure out whether response strategies are influential on customers' latter rating. It is worth mentioning that a customer might have visited the same hotel for numbers of times, however, only the ratings left after the previous negative rating are taken into consideration in this study. Supposing that a customer visited a hotel four times and gave ratings 4, 5, 3, and 4 out of 5, respectively. In this case, the customer's first negative rating is 3, so the second rating, which comes after the negative rating, is 4. Additionally, there might be more than one rows of data belonging to same customers of the same hotels. For instance, considering a customer who visited the same hotel for three times with the ratings of 2, 1, and 4, so two rows of data are emerging because the customer visited the same hotel twice after giving two negative ratings. As such, the first row belonging such customers includes the decrease from the ratings (i.e., $2 \rightarrow 1$), and the second row consists of the increase from the ratings (i.e., $1 \rightarrow 4$).

Moreover, the action-included category is divided into two categories as general action, included and specific action-included responses. General action-included responses refer to responses which are promising to take necessary actions in a generic way. For example,

- "Your comments have been passed to relevant department to be investigated further."
- "Thanks for your review, we will take necessary actions where needed."

Unlike general action-included responses, specific action-included responses contain specific solutions which will be provided or were provided when the customer was at the hotel. For instance,

- "We are extremely sorry for the Wi-Fi problems occurred during your stay. Please be assured that the engineering department sorted the problem out."
- "We are happy to provide another room during your stay."

The remaining categories are Admittance of Mistake(s), Explanatory, and Direct Contact Requested which are the same categories existing in the previous study already.

In this phase of study-1, responses to the repeated customers who stayed in London 5-star hotels are investigated. The total number of rows in the data is 432, including 343 responded reviews. The reason for selecting only London 5-star hotels is that response annotations are done manually in this phase of the first study due to the more diminutive size of the data than the data in the first phase of this study and the addition of new categories. However, analysing the comments of returning customers of hotels with different stars would reveal a more comprehensive outcome.

Table 12 shows the distribution of response contents based on rating difference (the difference between the ratings given by customers on their next visit and their previous negative ratings). Rating differences are in the range of –2, -1, 0, 1, 2, 3, and 4. In this range, 2 and -1 differences signify a decrease in the customers' next rating. In other words, customers' latter ratings are lower than their negative ratings given on their previous visit. Supposing that a customer's first negative rating is 3, and latter rating is 1, so the difference becomes -2. Likewise, if customers give the same ratings those are both negative, the difference becomes 0. The other rating differences [1, 2, 3, and 4] denote that there is an increase in customer's next rating; however, this does not mean the subsequent rating is certainly positive unless the difference is either 3 or 4. More specifically, if a customer's first rating is 10 and the subsequent is 2, there is an increase by 1 despite two negative ratings.

When looking at the table, there are total 196 responses for customers whose subsequent ratings increased by 1 or 2 compared to their first rating while there are 64 responses for customers whose rating increased by 3 or 4. The number of responses given to customers whose first and subsequent negative ratings are equal is 53 while the number of responses given to customers whose subsequent ratings decreased by 1 or 2 is 3.

Table 12 numerically and Figure 25 visually shows the distribution of response categories according to rating differences. Considering the responses given to the customers whose subsequent ratings increased by 4, which is the highest increase, 54.2% of them include specific action-included and 50% of them contain direct contact requested categories. Regarding the responses given the customers whose subsequent ratings increased by 3, strikingly 42.5% of them include specific action-included and 55% of them contain direct contact requested categories. Looking at the responses given to the customers whose subsequent ratings increased by 2, 44.2% of them include specific action-included, 37.2% of them contain direct contact requested categories, 33.6% of them include admittance of mistake(s) categories, and 32.7% include explanatory categories. In regard to the responses given the customers whose subsequent rating increased by 1, the distribution of categories seems balanced with a specific action, general action, and explanatory categories being between 32–36% while other categories being between 21.5–26.5%. Last, regarding the responses given to the customers whose subsequent ratings are equal to or less by 1 than their previous negative ratings, the percentages of explanatory and general action categories are greater than the percentages of specific action and direct contact categories.

In brief, when looking at the responses given to customers whose subsequent ratings are greater than previous ratings by 2, 3, and 4, the percent of specific action (marked as green on Figure 27) and direct contact requested (marked as orange on Figure 27) categories are the highest two categories among others.

Table 12 Distribution of Response Categories Based on Rating Difference

Rating Difference	Response Count	Admittance of Mistake (%)	Direct Contact Request (%)	Explanatory (%)	General Action (%)	Specific Action (%)
-2	4	0	50	25	25	0
-1	26	11.5	11.5	26.9	30.8	19.2
0	53	15.1	13.2	24.5	28.3	20.8
1	83	26.5	21.7	34.9	32.5	36.1
2	113	33.6	37.2	32.7	22.1	44.2
3	40	22.5	55	30	20	42.5
4	24	16.7	50	29.2	4.2	54.2

Figure 27 The Distribution of Response Categories Based on Rating Difference

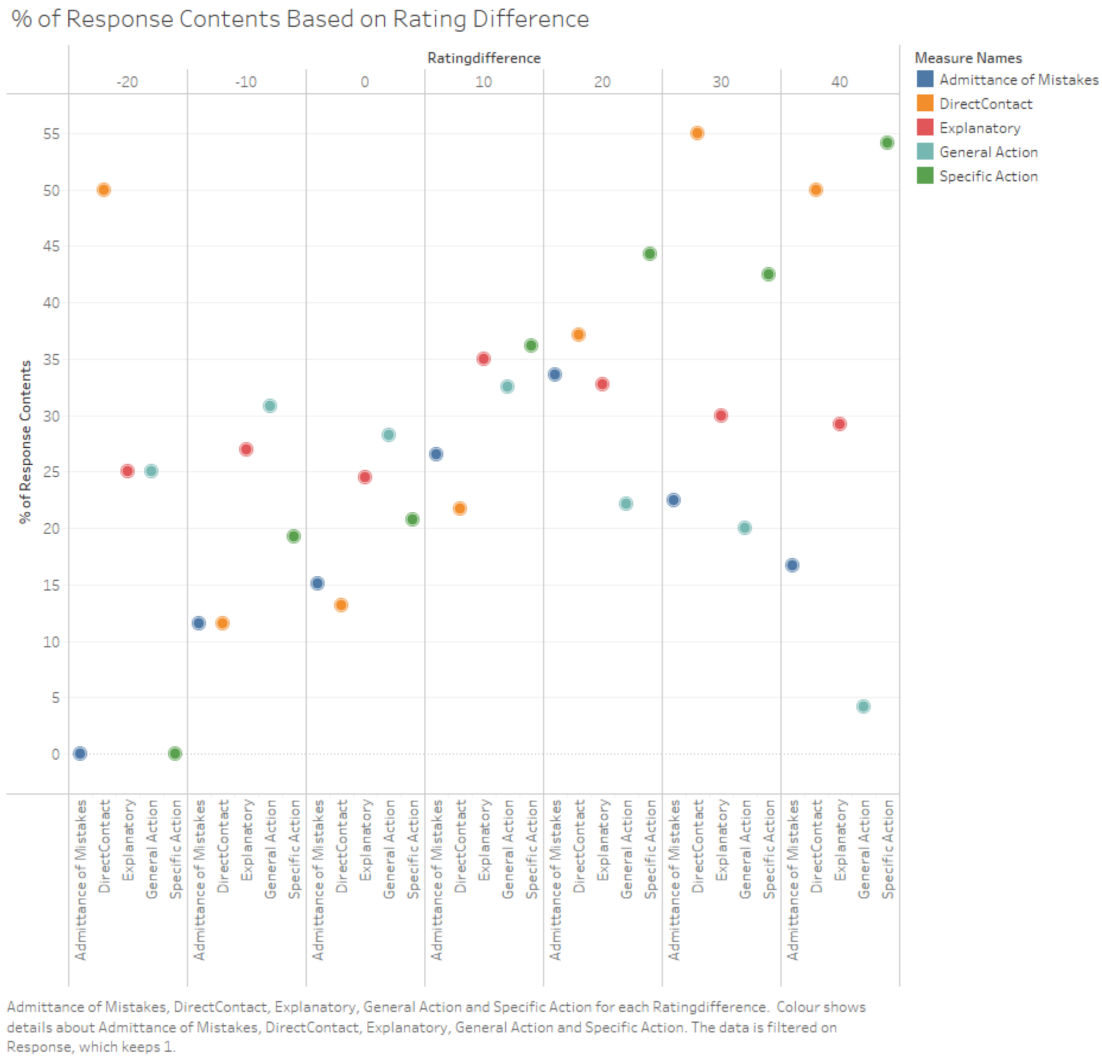


Table 13 numerically and Figure 28 visually shows the distribution of response categories based on the subsequent ratings. The number of customers whose ratings turned from negative to positive is 245 where 96 customers gave 4, and 149 customers gave 5 ratings. The number of customers who gave negative rating in their next visit as well is 96, where 55 customers gave 3, 26 customers gave 2, and 17 customers gave 1.

Regarding the responses given to customers whose subsequent ratings are the highest, which is 5, after the previous negative experience, it is clearly seen that the categories of direct contact request and specific action included are the most common categories existing in responses with 45% each. The distribution of admittance of mistake(s), explanatory and general action categories are 30.9%, 28.2%, and 20.1%, respectively.

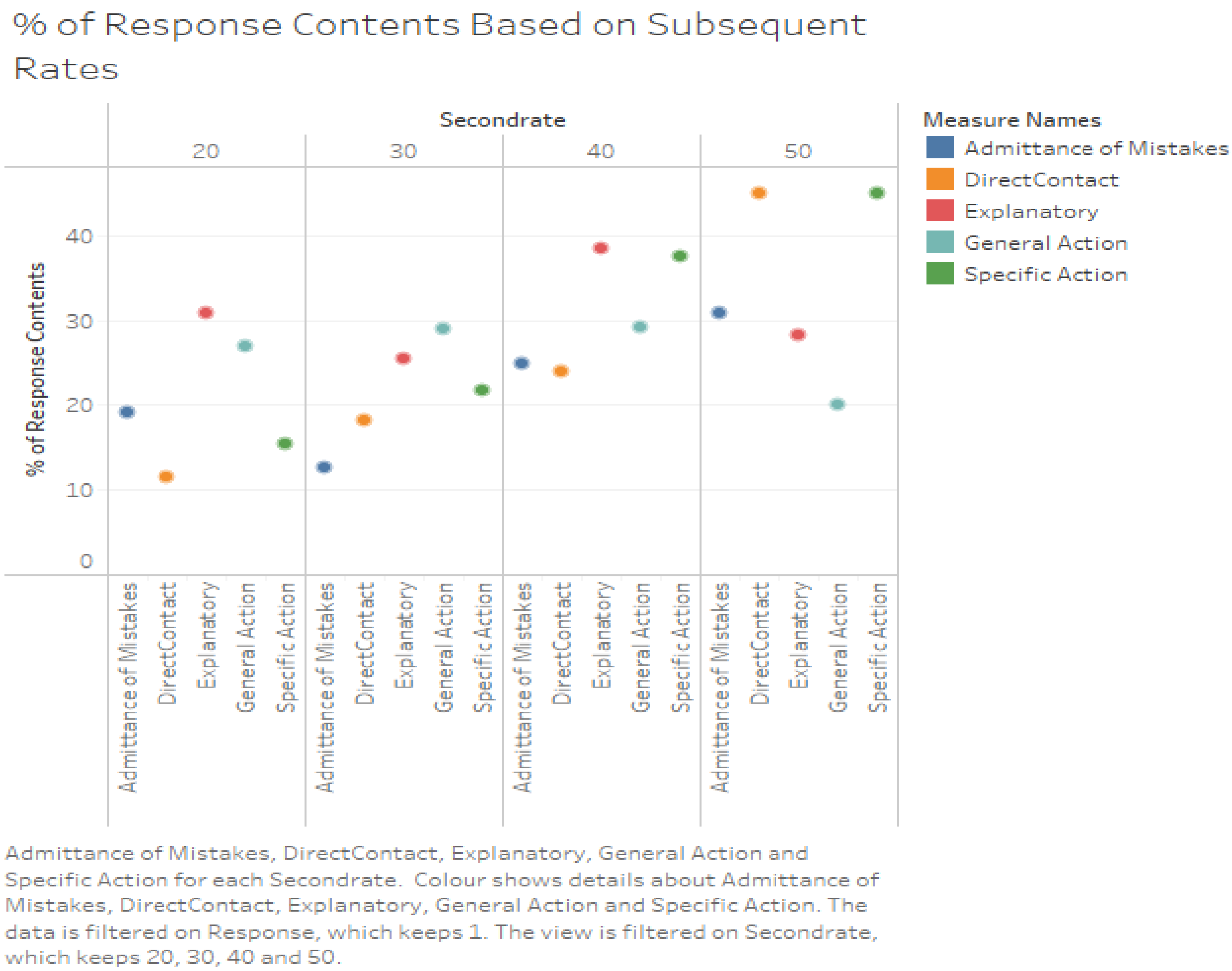
Looking at the responses given to customers whose subsequent ratings are 4, the category with the highest distribution is explanatory with 38.5%, followed by specific action category with 37.5%. The distributions of direct contact requested, admittance of mistake(s), and general action categories are 24%, 25%, and 29.2%, respectively.

In regard to the responses given to customers whose subsequent ratings stayed negative, there is a considerable difference in the percentile of specific action, direct contact requested, and admittance of mistakes categories compared to the responses provided customers whose subsequent ratings are positive; 21.8 % of responses involve specific action category, 18.2% involve direct contact requested category, and 12.7% contain admittance of mistake(s) category, which was given to customers whose subsequent ratings are 3. Likewise, 15.4% of responses involve specific action category, 11.5% involve direct contact requested category, and 19.2% contain admittance of mistake(s) category, which were given to customers whose subsequent ratings are 2.

Table 13 The Distribution of Response Contents Based on Subsequent Ratings

Second Rate	Response Count	Admittance of Mistake (%)	Direct Contact Request (%)	Explanatory (%)	General Action (%)	Specific Action (%)
2	26	19.2	11.5	30.8	26.9	15.4
3	55	12.7	18.2	25.5	29.1	21.8
4	96	25.0	24	38.5	29.2	37.5
5	149	30.9	45	28.2	20.1	45

Figure 28 The Distribution of Response Contents Based on Subsequent Ratings



4.4.1 A Logistic Regression Analysis of Subsequent Customers' Ratings

To observe subsequent ratings of customers, who visited the same hotel after leaving a negative review, a logistic regression analysis was conducted. The main purpose of this analysis is to understand whether there is an increase in subsequent ratings of customers, considering the independent variables that are response categories (admittance, direct contact requested, general action inclusive, specific or personalized action inclusive, and explanatory), response length, and response days (response speed). The dependent variable is increased which is 1 if repeated customers' subsequent ratings are greater than their previous visit with negative valence. The dependent variable is 0 if repeated customers' subsequent ratings are not greater (lower or equal) that their previous visit with negative valence.

Table 14 shows the omnibus test of model coefficients on chi-square test that implies that the overall model is predictive of increase in subsequent ratings. The significance value is less than 0.001 (chi-square = 47.854, degree of freedom = 7), which indicates that the current model outperforms the null model.

Table 14 Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step1	Step	47.854	7	$< .001$
	Block	47.854	7	$< .001$
	Model	47.854	7	$< .001$

The model summary (shown on Table 15) provides the -2 Log likelihood (–2LL) and pseudo–R square values for the full model. The R square values Cox&Snell, and Nagelkerke suggest that the model explains roughly 24% and 32% of the variation in the outcome, respectively.

Table 15 Model Summary

	-2 Log likelihood	Cox&Snell R Square	Nagelkerke R Square
Step1	190.588^{a}	0.243	0.324

The Hosmer & Lemeshow test (see Table 16) of the goodness of fit suggests the model is a good fit the data as p = 0.361 (> 0.05).

Table 16 Hosmer and Lemeshow Test

	Chi-square	df	Sig.
Step1	8.779	8	0.361

Table 17 shows the classification results of the model. Accordingly, 59 of 86 instances were correctly classified to be no increase at 69%. 61 of 86 instances were accurately classified to be increased at 71%. Eventually, the overall accuracy percentage of the model is 70%.

Table 17 Classification Table

	Observed		Increase 0	1	Percentage Correct
Step1	Increase	0	59	27	69
		1	25	61	71
	Overall Percentage				70

Table 18 shown below provides the variables in the equation with the columns of regression coefficient (B), standard error (S.E.), Wald statistic, degree of freedom (df), significance (Sig.), and odds ratio (Exp B) for each variable category. Looking at the results, the b coefficients are positive for each response category and negative for response length and response speed. Admittance of inclusive responses, direct contact request inclusive responses, and specific action inclusive responses have significant relationships (p-values are all below 0.05) with increase in subsequent rating. There is no strong enough relationship between other independent variables and increases in subsequent rating.

Because the logistic regression conducts the analysis on the log odds, it is worth it to interpret the odds ratio (Exp (B)) rather than regression coefficient (B). When looking at the odd ratios on the table, the odds of increase in subsequent rating is 5.928 times greater for providing the response including direct contact request as opposed to providing response without direct contact request. Similarly, the odds of increase in subsequent rating is 5.293 times greater for providing the specific action inclusive response as opposed to providing a response that is lacking specific action. In addition, the odds of increase in subsequent rating is 5.286 times greater for providing the admittance inclusive response as opposed to providing responses without an admittance inclusive response.

Table 18 Variables in Equation

							95% C.I for EXP (B)	
	B	S.E	Wald	df	Sig.	Exp(B)	Lower	Upper
General Action	.264	.480	0.302	1	.582	1.302	.508	3.335
Admittance	1.665	.511	10.600	1	**.001**	5.286	1.940	14.404
Explanatory	.776	.458	2.795	1	.095	2.151	.876	5.277
Direct-Contact	1.780	.464	14.717	1	**.001**	5.928	2.388	14.717
Specific Action	1.666	.426	15.328	1	**.001**	5.293	2.298	12.191
Response Length	-.002	.001	5.572	1	.018	.998	.997	1.000
Response Speed	-.002	.003	.455	1	.500	.998	.992	1.004
Constant	-.617	.417	2.189	1	.139	.540		

4.5 Discussions and Theoretical Implications

The proliferation of user generated content has led firm engagement to gain more importance to manage communications with customers. For online reviews, management response seems to be an intervention tool when interacting with customers (Ma et al., 2015). Considering that a service failure can cause customer dissatisfaction and thus proliferation of negative eWOM, managers seek strategic ways to respond to complaints to mitigate detrimental impacts. These strategies could be adjusted based on quantitative and qualitative attributes of responses. Even though existing research demonstrates the impact of quantitative attributes of responses (e.g., response rate, speed, length) on hotel performance, customer engagement, and customer ratings, the quantitative attributes do not reflect the heterogeneous nature of management responses (Li et al., 2020), and therefore response content is significant to understand how strategically managers act in case of service failure. Thus, this study analyses response strategies based on both qualitative and quantitative attributes.

Qualitative paradigm research has explored the impact of the content based on accommodativeness vs. defensiveness, generic vs. specific, response sentiment, showing empathy, explanation vs. meta-discourse on customers' attitude, brand evaluation, purchase intention, and trust. Nevertheless, much of the current literature overlooks a detailed analysis of what content strategies hotels develop and the differences between response strategies based on hotel class and overall customer rating. With the help of natural language processing and deep learning techniques, this study presents a comprehensive classification of management response content to negative reviews and enhance our understanding of managerial reactions in case a service failure. The first phase of this study adds weight to this stream by specifically categorizing response content into five strategic moves: brand positioning, admittance inclusive, action inclusive, explanatory, and direct contact request inclusive and reveals their correlation with the overall rating of customers. The results present a positive correlation between overall hotel rating and brand positioning, admittance inclusive, and direct contact request inclusive response strategies. Although it is not possible to talk about the impacts based on a correlational analysis, it is worth it to consider formulating these strategies. Additionally, hotels with a higher overall rating than average adopt each response category more often than other hotels do. It is unclear whether these response strategies have an impact on the average rating considering that there are many factors influencing hotel ratings or whether high-rated hotels pay more attention to response strategies. Hence, it would be interesting to investigate the effect of these content strategies on subsequent ratings in future research.

In addition, limited research has provided evidence on the efficacy of different response strategies to returning customers. Although it has been demonstrated that managerial responses have positive impact on customers' future ratings (Gu & Ye, 2014; Proserpio & Zervas, 2017; Sheng et al., 2019), it has not been widely examined what specific content strategies are influential in future ratings of dissatisfied returning customers. The second phase of this study reveals response strategies that are likely to increase subsequent rating of repeated customers who had a negative service experience in their previous visit. Unlike the response classification approach adopted in the first phase, the action category is analysed in two parts: specific action that includes an action statement about a certain problem partaking in corresponding review and general action that contains a superficial declaration implying necessary actions will be taken by organization. The findings demonstrate that referring to specific action inclusive statements are more likely to increase subsequent ratings of repeated customers. This suggests that customers may want to see customized solutions in attached responses rather than generic statements. Second, responses that include statements of taking responsibility for a service failure appear to be more effective than those that do not in future ratings. Third, making a direct contact request via email and phone in response seems to be effective in improving the next rating of repeated customers. This implies that customers may want to be in contact with the response provider for their complaints about a more customized solution. However, the findings indicate that response speed has no impact on future ratings of dissatisfied customers. This finding is in line with Min et al. (2015) and Sheng et al. (2019) who also find no impact of response speed on ratings. Interestingly, response length is also not influential in subsequent ratings of dissatisfied customers. This may indicate that the content of the response is more important to dissatisfied customers, regardless of how long the response text is. As discussed earlier, the quantitative characteristics of the response alone are not sufficient to have efficient communication with customers. Strategically formulated response content can play a significant role in managing relationships with customers. Eventually, this study presents several comparative and descriptive analysis of management responses to negative reviews, thus, contributing to the management responses literature and academic understanding of effective service recovery in an online context.

An additional contribution of this study arises from big data analytics from social media. With the rapidly growing firm and consumer activities, social media has become a valuable source of big data (Kunz et al., 2017), supporting the marketing activities of many firms (Hajli & Laroche, 2019). Unlike traditional methods such as survey and focus groups, big data analytics from social media allows us to monitor and analyse a vast amount of data to generate knowledge and gain useful insights. This study supports current research on the knowledge-based view of a firm (Grant, 1996; Nonaka, 1994; Sheng et al., 2019) by investigating how organizations leverage the vast and various type of data from social media to adopt new ways of responding to dissatisfied customers. A large amount of text, which would take a lot of time to read and label with human effort, was classified adopting machine learning and deep learning algorithms. The quantified textual data was then combined with other quantitative data to provide comprehensive analysis of the impact of firms' strategies to consumer reviews. This contributes to the social media analytics research, especially in marketing and strategic management, in the way of adding value to our knowledge of data driven strategies and providing applicable approaches for future research in this scope.

Another contribution of this research is to present the capability of different strategies for firm engagement in social networking to improve social media marketing. Despite the fact that social media use has become pervasive by both users and marketers, research on the strategic use of social networking sites is still evolving. This research examines firms' engagement in an online review platform and contributes to the extensive social media marketing literature with the exploration of strategies of firm-generated content. Particularly, the results of this research support IT-enabled dynamic capabilities that underlines the firms' dynamic capabilities that are obtained or improved with the help of IT (Mikalef & Pateli, 2017). From the marketing perspective, IT has been considered as a major enabler of dynamic capabilities to exploit competitive advantage through service excellence and customer intimacy. This research present evidence of improved customer relationship management by investigating effective communication styles of firms interacting with customers. This contributes to the marketing view of dynamic capabilities with emphasis on social media use and marketing strategies to enhance the customer relationship management capabilities of the firms (Wang & Kim, 2017).

CHAPTER 5:

STUDY–2: AN INVESTIGATION OF THE EFFECT OF MULTITOPIC MATCHING BETWEEN RESPONSES AND REVIEWS ON RATING GROWTH

5.1 Study Background

Personalisation is a competitive advantage strategy that includes learning, matching, and delivering products and services to customers. Customers gain from personalisation because it reduces confusion by highlighting solutions that satisfy their needs (Murthi & Sarkar, 2003). Personalisation seeks to increase customer pleasure by enhancing the quality of decisions, hence fostering brand loyalty. If handled with care, personalised marketing can increase client value by broadening the relationship(Peppers & Rogers, 1993). Blom and Monk (2003) defined personalisation as a process that modifies functionality, content or distinctiveness, interface, and information access based on the individual customer's relevance. Aksoy et al. (2021) also claim that the concept of personalisation also includes displaying and utilising consumer data to provide a personalised customer experience based on their needs and interests.

The importance of tailoring managerial response content to customer reviews has been shown in several studies. Personalized management response carries additional information for potential customers (Whang & Chaudhry, 2018) and could refer to attitudinal consensus between manager and customer (Zhang et al., 2020). As opposed to a standardized response, which tends to be less relevant to the review, a personalized or specific response that tackles a problem or shows empathy plays a pivotal role in gaining prospective customers' trust and cultivating the perceived quality of communication (Min et al., 2015).

A topic-matched response allows managers to transmit more relevant information that meets customers' functional needs (Garaus et al., 2015; Kopp et al., 2018). As for the positive reviews, managers can focus on the specific topic and talk with more details about corresponding advantages that customers are satisfied with. That is, they can improve potential customers' perception of commendable aspects of hotel and the intention to purchase (Zhang et al., 2020). With regard to negative reviews, responses that include why the issues occurred or how to solve them can strengthen customers' perception of hotel's attention (Zhang et al., 2020). In addition, topic matching between response and review can improve the perceived effectiveness and reduce the potential biases for customers (Lee

& Cranage, 2014; Li et al., 2017; Wei et al., 2013). In Zhang et al. (2020), topic matching has a positive effect on hotel rating increases. To calculate the topic matching degree, they first classify customer reviews with a multiclass text classification approach and then find the probability that response and corresponding review include the same topic. However, in this approach, they consider that a review and a response can only include one topic. In fact, customers mostly mention more than one topic in their reviews. That is because this study extends the study of Zhang et al. (2020) by applying the multilabel classification approach. This allows that more than one topic can be assigned to reviews and responses. Differently from Study–1 in this research, where responses to only negative reviews are investigated, this study investigates the effect of management responses to both positive and negative reviews of customers.

5.2 Methods

5.2.1 Data

The data are collected from randomly selected 3-, 4-, and 5-star London hotels for the period of January 2016 to December 2019. London was chosen because of being a well-known travel destination on a global scale and the proliferation of the number of English reviews. The feedback from guests and management on a variety of hotels in the city satisfies the criteria for sampling and research purposes. In addition, concentrating on a particular market is valid since businesses in the same area are regionally localised and engaged in regional market competition. Companies operating in a market can learn how to achieve competitive advantages by researching customer feedback and the management styles of businesses in that market. The total number of hotels is 105, with being 35 hotels for each hotel grade. While the number of customer reviews is 237,560, 183,401 of which were responded to by managers. That means the response ratio is approximately 77% in the whole data set.

5.2.2 Review Topics

The topics of customer reviews are defined according to Zhang et al. (2020). They use high-frequency feature words analysis to epitomize representative topics partaken in customer reviews. They tokenized almost 220, 000 customer reviews and obtained a list of top 200 high-frequency feature words to determine the representative topics. As a result, they come up with five topics according to the attribute and meaning of these words: food and beverage, price, facilities and amenities, service, and environment. The food and beverage topic includes comments regarding food, drink, and dining experience. The topic of price involves customers' opinions about hotel's pricing such as parking fee, accommodation cost, and any other related fees. The facilities and amenities topic consists of customers' essential lodging experience about places, equipment's, comfort, convenience, and so on. The topic of service represents service attitude, quality, and responsibility of hotel staff. Lastly, the environment topic involves customers' thoughts about the convenience of the hotel's location such as places around the hotel or distance from central, airport, and shopping malls. As opposed to the majority of earlier research that only considered lodging issues that were pre-determined by travel websites, such as room cleanliness and sleep quality (Zhang et al., 2011), or an overwhelming amount of scattered and meticulous topics (Guo et al., 2017), the five subjects defined by Zhang et al. (2019) are the typical "themes" that emerged from the consumer review corpus as a whole.

5.2.3 Data Annotation and Classification

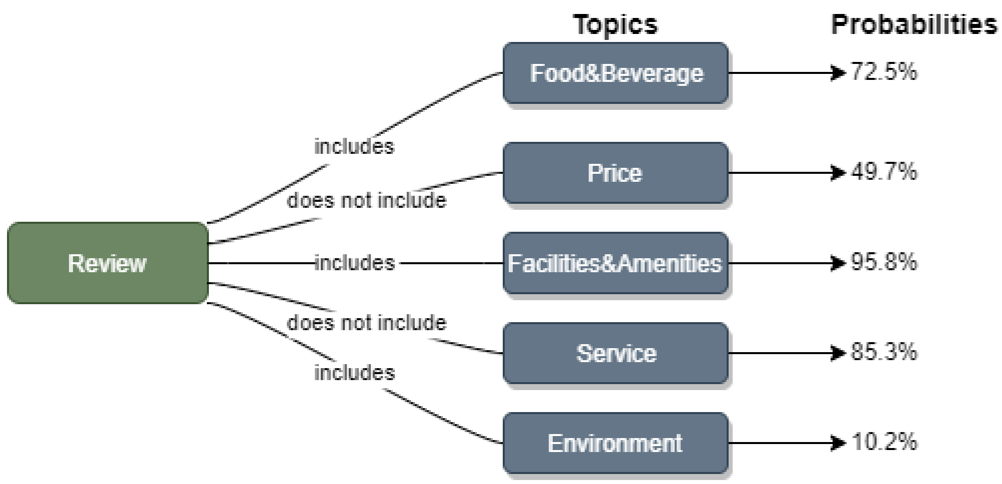
To annotate the text data, 2,000 reviews and 2,000 management responses are randomly selected from the data set. Two distinct data sets are created to label review data sample and response data sample separately, which approach is different from the study of Zhang et al. (2020) as they classified management responses based on the review data classification model. The reason this research involves separate a data annotation for review and response data is that the text classification model of the review sample data does not produce high accuracy on response sample data. This might be caused by the wording difference between managerial response and customer review. Eventually, 2,000 reviews and 2,000 responses are annotated. Apart from that, a crucial contribution of this study is conducting a multilabel data annotation approach as opposed to the study of Zhang et al. (2020) who adopt the multiclass data annotation and classification approach. In the multiclass classification method, a sample can be assigned to one and only one label among more than two labels. A vital handicap of this approach is to assign a customer review to only one topic among five topics. However, most customer reviews can involve more than one topic. For instance, a customer can voice out a satisfaction of food and beverage and a dissatisfaction of price. In this case, the review contains more than one topic. To overcome this barrier, this study conducted a multilabel data annotation approach where a document can be assigned to a set of target labels. That provides a customer review to belong to more than one topic at the same time or none.

This study uses Prodigy as an annotation tool due to its allowance and efficiency for the multilabel annotation approach. spaCy's multilabel text categorizer is adopted to classify review and response data, which comes as a built-in feature in Prodigy tool. The text categorizer of Spacy uses a simple CNN architecture in its pipeline, allowing users to save time in modelling text classification algorithm with deep learning. The annotated review data are split into training and test sets with the 0.2 ratio, which means that 80% of annotated data are training set and 20% of annotated data are test set. Eventually, the training data contains 1600 customer reviews while the test data contains 400 samples. A multilabel text classification model then is generated with training data and this model achieves the 87% classification accuracy on the test set. The same steps are conducted on the annotated response data where the model reaches 83% classification accuracy on the test set of response data. These accuracy results are higher than the achieved accuracy in the study of Zhang et al., where they achieve approximately 70% classification accuracy on review data using support vector machines.

5.2.4 Topic Matching Degree

The classification models obtained from review and response data samples are applied in the main data set, and thus the probabilities of which topics the relevant reviews and responses belong to are calculated. It is considered in this study that reviews and responses include topics whose probabilities are higher than 70%. Figure 29, as an example, shows a customer review with topic probabilities as 72.5% for food and beverage, 49.7% for price, 95.8% for facilities and amenities, 85.3% for service, and 10.2% for environment. In this case, this customer review includes the topics food and beverage, facilities and amenities, and service but does not include the topics price and environment.

Figure 29 Assigning Topics to Review

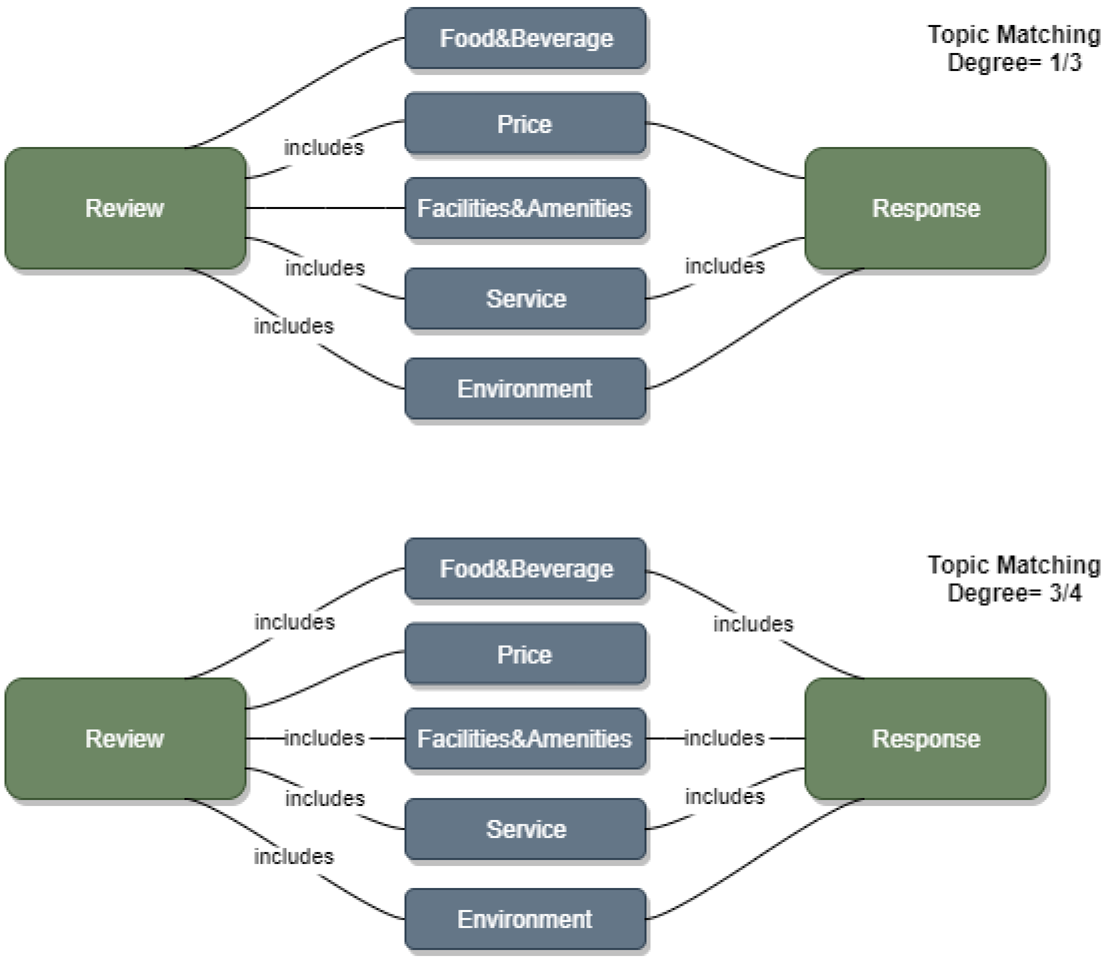


After assigning topics to reviews and responses, topic matching degrees are calculated by looking at how many of the topics mentioned on relevant customer reviews are also included in relevant management responses. Thus, the topic matching degree formula is shown as follows:

$$TopicMatching = \frac{The\ number\ of\ topics\ in\ response\ that\ mathcing\ the\ relevant\ review}{The\ number\ of\ topics\ in\ review}$$

Figure 30 shows an instance of this process. A customer review contains the topics price, service, and environment while corresponding response includes only service from these topics. Therefore, the topic matching degree of this response is 1/3, which is approximately 33.3%. Another review consists of all the topics except price whereas the corresponding response includes three of these topics, which means that topic matching degree is 3/4 which is 75%.

Figure 30 Calculating Topic Matching Degree Between Response and Review



5.2.5 Variables and Model Specification

In this study, daily reviews and responses are aggregated to monthly level (*t*) in the form of a hotel-month panel. Table 19 indicates a description of variables. The dependent variable is the growth of online customer ratings (*RatingGrowth*) measured as growth of average rating from month t-1 to month t, which is specified as follows:

$$RatingGrowth = \frac{ReviewValence_{i,t} - ReviewValence_{i,t-1}}{ReviewValence_{i,t-1}}$$

While investigating the effect of topic matching degree, this study controls other common characteristics of responses and reviews. RatingGrowth is modelled as a function of the two groups of variables. The first group consists of descriptors of management response, which are the average topic matching degree between customer review and management response TopicMatching, ResponseRatio, the average text length of response ResponseLength, and the average response days ResponseSpeed of a hotel *i* in period *t–1*. Because prior studies demonstrate that eWOM generation is a dynamic process that can be evaluated by valence, variance, and volume that influences subsequent ratings (Guo & Zhou, 2016; Moe & Schweidel, 2012; Wu & Huberman, 2008), the average ratings of customers ReviewValence (the average of 1-5 star ratings given by customers after consumption), the variance of reviews ReviewVarience (the fluctuation range of customer ratings), and the number of reviews ReviewVolume (the total number of customer reviews) of a hotel *i* in period *t–1* are added in the second group as descriptors of online customer reviews.

Adding sales data would have made the model informative of hotels' performances. Besides, analysing the impact of some variables such as the career positions of response providers, price information of hotels, and demographic information of customers would reveal intriguing results.

Table 19 Description of the Variables

Variables	Description
Dependent Variables	
$RatingGrowth_{i,t}$	Average rating growth rate from month t-1 to month t for hotel i
Independent Variables	
$TopicMatching_{i,t}$	Average topic matching degree for hotel i in month t
$ReviewValence_{i,t}$	Average rating of reviews for hotel i in month t
$ReviewVariance_{i,t}$	Variance of review ratings for hotel i in month t
$ReviewVolume_{i,t}$	Total number of reviews for hotel i in month t
Control Variables	
$ResponseRatio_{i,t}$	The number of management responses ratio to total number of reviews for hotel i in month t
$ResponseLength_{i,t}$	Average length of management response texts for hotel i in month t
$ResponseSpeed_{i,t}$	Average response time (in days) to reviews for hotel i in month t

The descriptive statistics of all the variables is shown on Table 20. Amongst these variables, standard deviations of ResponseLength, ResponseSpeed, and ReviewVolume are substantially greater than other variables in model. Therefore, logarithmic values of these three variables are taken to obtain close-to-normal distributions to make overall data more consistent. Considering that there is mostly a time lag between management response, customer review, and eWOM effects are continuous for weeks (Xie et al., 2014), lagged variables are used to investigate the effect of management responses on subsequent ratings. Reviews' datetime stamps are used in the analysis, so the response variables stand for hotel *i*'s responses to time *t–1's* reviews. As for the review variables, average review valence, average review volume, and variance of ratings of hotel i in the previous *t–1* periods are computed. That is because the aggregated values shown on the website might affect ratings in time t.

Table 20 Descriptive Statistics of Variables

	Mean	Std. Dev.	Min	Max
RatingGrowth	.0081178	.1435822	-.75	3
TopicMatching	.4193294	.1967642	0	1
ResponseRatio	.8319567	.3157921	0	1
ResponseLength	469.777	200.6321	98.6	3701
LnResponseLength	6.080672	.3719413	4.591071	8.216358
ResponseSpeed	13.22274	58.84211	0	1521

LnResponseSpeed	1.501758	1.069116	-.7472144	7.327123
ReviewValence	4.139403	.557003	1	5
ReviewVariance	.9875235	.6899828	0	8
ReviewVolume	28.37452	19.90673	1	118
LnReviewVolume	3.054078	.842235	0	4.770685

Additionally, the variance inflation factor (VIF) is controlled to extract multicollinearity issues and the results indicate that all VIF scores are below 2 (see Table 21), which is less than the common threshold. As such, multicollinearity is not a problem in the model. The correlation coefficients among all variables are relatively small (see Table 22), which also means the multicollinearity problem of variables was not an issue in the estimation model.

Table 21 The Variance Inflation Factor (VIF)

Variable	VIF	Tolerance
TopicMatching	1.34	0.7465
ResponseRatio	1.03	0.9731
LnResponseLength	1.38	0.7269
LnResponseSpeed	1.03	0.9676
ReviewValence	1.87	0.5345
ReviewVariance	1.74	0.5739
LnReviewVolume	1.14	0.8794
Mean VIF	1.36	

Table 22 Correlation Matrix of Variables

	1	2	3	4	5	6	7	8
1.RatingGrowth	1							
2.TopicMatching	.0529	1						
3.ResponseRatio	.0119	.0574	1					
4.LnResponseLength	.1170	.4948	-.1559	1				
5.LnResponseSpeed	-.0243	-.0543	-.0193	-.0344	1			
6.ReviewValence	-.4297	.0279	.0553	-.0914	-.0591	1		
7.ReviewVariance	.2914	-.0181	.0439	.0867	.0142	-0.6508	1	
8.LnReviewVolume	.1042	.0525	.0245	-.0831	-.1639	.3036	-0.1680	1

The data set is hierarchical; more specifically, the hotel-month observations have a nested structure. Customer reviews and management responses of hotel i in month t are clustered for each hotel i as a level 1, which is the identifier of level 2. It is worth to note that there may be imperceptible heterogeneity issues at the hotel level such as severe service crisis, major change in service quality, and managerial expertise and potential time effects changing along the hotels that can affect ratings. To consider this issue and analyse the clustered data, a mixed effect multilevel model approach is adopted which is inclusive of both fixed and random effects and takes account of deviation between groups and estimation of group effects. The estimation model is specified as follows:

$$
\begin{aligned}
RatingGrowth_{i,t} \\
&= \beta_0 + \beta_1 TopicMatching_{i,t-1} + \beta_2 ResponseRatio_{i,t-1} \\
&+ \beta_3 LnResponseLength_{i,t-1} + \beta_4 LnResponseSpeed_{i,t-1} \\
&+ \beta_5 ReviewValence_{i,t-1} + \beta_6 ReviewVariance_{i,t-1} \\
&+ \beta_7 LnReviewVolume_{i,t-1} + \tau_i + u_{0i} + u_{it} + e_{it}
\end{aligned}
$$

The moderating effects of the eWOM and hotel star on the topic matching degree are further investigated. Accordingly, the equation with the interaction terms is designated as follows:

$$
\begin{aligned}
RatingGrowth_{i,t} \\
&= \beta_0 + \beta_1 TopicMatching_{i,t-1} + \beta_2 ResponseRatio_{i,t-1} \\
&+ \beta_3 LnResponseLength_{i,t-1} + \beta_4 LnResponseSpeed_{i,t-1} \\
&+ \beta_5 ReviewValence_{i,t-1} + \beta_6 ReviewVariance_{i,t-1} \\
&+ \beta_7 LnReviewVolume_{i,t-1} + \beta_8 TopicMatching_{i,t-1} \\
&* ReviewValence_{i,t-1} + \beta_9 TopicMatching_{i,t-1} * ReviewVariance_{i,t-1} \\
&+ \beta_{10} TopicMatching_{i,t-1} * LnReviewVolume_{i,t-1} \\
&+ \beta_{11} TopicMatching_{i,t-1} * HotelStar_i + \tau_i + u_{0i} + u_{it} + e_{it}
\end{aligned}
$$

5.3 Empirical Results

Table 23 shows the estimation results for the response and review data using a multilevel modelling approach. Model 1 results show that topic matching (TopicMatching) has a significant positive influence on the subsequent ratings (coeff. = 0.0252, $p < 0.05$), suggesting that the higher level of topic matching of management response, the higher hotel rating increase in the following month. The response ratio (ResponseRatio) coefficient is positive and statistically significant (coeff. = 0.0216, $p < 0.05$), signifying that the higher response rate provides the higher rating growth in the subsequent period. Likewise, the response length coefficient is positive and statistically significant (coeff. = 0.0264, $p < 0.01$), suggesting that the longer response text has an influence on future ratings in the monthly period. Moreover, it is observed that response speed has no statistical significance in rating growth in the next month.

Table 23 Estimation Results by Month

RatingGrowth	Model 1	Model 2
Fixed Effects		
TopicMatching	$.0252^{**}$	$.6288^{***}$
	(.0119)	(.0886)
ResponseRatio	$.0216^{**}$	$.0211^{**}$
	(.0101)	(.0099)
LnResponseLength	$.0264^{***}$	$.0293^{***}$
	(.0074)	(.0073)
LnResponseSpeed	$-.0023$	$-.0032^{*}$
	(.0018)	(.0018)
ReviewValence	$-.2413^{***}$	$-.1702^{***}$
	(.0062)	(.0111)
ReviewVariance	$-.0112^{***}$	$.0131$
	(.0034)	(.0066)
LnReviewVolume	$.0101^{***}$	$.0286^{***}$
	(.0039)	(.0063)
TopicMatching*ReviewValence		$-.1520^{***}$
		(.0207)
TopicMatching*ReviewVariance		$-.0516^{***}$
		(.0127)
TopicMatching*LnReviewVolume		$-.0447^{***}$
		(.0127)

TopicMatching*HotelStar		$.0498^{***}$
		(.0133)
Constant	$.8010^{***}$	$.4157^{***}$
	(.0567)	(.0702)
Random Effect St. Deviations		
Hotel Level	$.0907^{***}$	$.0846^{***}$
	(.0069)	(.0066)
Month Level	$.0161^{***}$	$.0134^{***}$
	(.0024)	(.0024)
Month Time Effect	$.0000^{***}$	$.0000^{***}$
	(.0000)	(.0000)
N	3421	3421
LogLikelihood	3091.9172	3134.6453
Chi-square	701.87	678.30

Note: The independent variables of response and review are one month lagged. All estimations have robust error terms clustered at the hotel level. Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

Regarding review variables, results show that review valence (ReviewValence) has a negative and statistically significant effect on the rating growth (coeff. = -0.2413, $p < 0.01$). This means that the higher average of ratings slows down the rating growth in subsequent period. This is reasonably expected and can be explained with the following example: an increase in the average rating of customers from 3 to 4.5 in the subsequent month equals 50% rating growth, whereas the average rating increase from 4 to 4.5 in the following month means 12.5 % rating growth. Likewise, review variability has a significant negative effect on the rating growth, indicating that customers exhibit negative attitude towards larger fluctuations in ratings (coeff. = -0.0112, $p < 0.01$). This might suggest that higher degree of variability in ratings can cause uncertainty about the service quality, which can affect customer decisions negatively while the rating process. By contrast, the coefficient of average volume of reviews (LnReviewVolume) is positive and significant (coeff. = 0.0101, $p < 0.01$). The higher customer engagement in previous *t-1* period, the higher rating growth in period *t*.

In addition, Model 2 further estimates the results of interaction terms of review variables and hotel star on topic matching. The results show that the moderating effect of ReviewValence (coeff. = -0.1520, $p < 0.01$) on topic matching is negative and significant. As such, the effect of topic matching response on hotel rating increase is weaker for hotels with high online review valence than hotels with low online review valence. The outcome can be explained by the fact that positive information typically signals greater diagnostic and expectation, which may encourage customers to focus more on service contrast and lessen the impact of management reactions. Additionally, ReviewVariance negatively moderates the influence of topic matching on hotel rating increase (coeff. = -0.0516, $p < 0.01$). When consumers confront a higher rating variance, they may become entangled in relatively negative reviews, and the efficacy of topic-matched management responses on customer satisfaction becomes less influential. Lastly, the moderating effect of LnReviewVolume on topic matching is also negative and significant (coeff. = -0.0447, $p < 0.01$). This might suggest that high customer engagement may lead customers to focus more on customer reviews while the rating process and the influence of personalized management responses become less efficient. However, there is a positive and significant moderating effect of hotel star on topic matching degree (coeff. = 0.0498, $p < 0.01$). This suggests that the influence of topic matching of management response in rating growth is stronger in higher grade hotels.

5.4 Robustness Check

Main empirical results shown above are subject to monthly review and response data. However, the median value of ResponseSpeed in the data set is 6 days, indicating that response is commonly displayed and so observable within a week. For this reason, a narrower time scale is adopted to observe the effect of managerial responses on weekly ratings. Table 24 presents the estimation results in a weekly time frame. The positive significant effect of topic matching and response ratio remain consistent with the base results. In addition, it is interesting that the effect of the length of response text is not significant while the effect of response speed is significant and negative on rating growth in a weekly period. This suggests that a longer response in the monthly period and a faster response in the weekly period have a positive effect on the increase of the customer ratings. As for the review variables, the negative significant effect of averaged rating of reviews and review variability on rating growth are coherent with base results, whereas review volume does not have significant effect on rating growth in the weekly period.

Furthermore, the moderating effects of averaged rating of reviews, review variability, and hotel star on topic matching remain consistent with the base results while no moderating effect is discovered for review volume on topic matching in a weekly period.

Table 24 Estimation Results by Week

RatingGrowth	Model 1	Model 2
Fixed Effects		
TopicMatching	$.0243^{***}$	$.394^{***}$
	(.0084)	(.0642)
ResponseRatio	$.0359^{***}$	$.0370^{***}$
	(.0099)	(.0099)
LnResponseLength	$.0014$	$.0012$
	(.0054)	(.0054)
LnResponseSpeed	$-.0036^{**}$	$-.0039^{***}$
	(.0015)	(.0015)
ReviewValence	$-.2962^{***}$	$-.2533^{***}$
	(.0035)	(.0062)
ReviewVariance	$-.01470^{***}$	$.0022$
	(.0017)	(.0032)
LnReviewVolume	$-.0025$	$.0030$
	(.0030)	(.0050)
TopicMatching*ReviewValence		$-.1078^{***}$
		(.0129)
TopicMatching*ReviewVariance		$-.0429^{***}$
		(.0070)
TopicMatching*LnReviewVolume		$-.0143$
		(.0109)
TopicMatching*HotelStar		$.0353^{***}$
		(.0103)
Constant	1.222^{***}	1.018^{***}
	(.0413)	(.0473)

Random Effect St. Deviations		
Hotel Level	$.1138^{***}$	$.1092^{***}$
	(.0083)	(.0080)
Month Level	$.0163^{***}$	$.0161^{***}$
	(.0020)	(.0020)
Month Time Effect	$.0000^{***}$	$.0000^{***}$
	(.0000)	(.0000)
N	3421	3421
LogLikelihood	3091.9172	3134.6453
Chi-square	701.87	678.30

Note: The independent variables of response and review are one month lagged. All estimations have robust error terms clustered at the hotel level. Standard errors in parentheses. * p < .1, ** p < .05, *** p < .01

5.5 Discussions and Theoretical Implications

Studies suggest that a convenient management response should be pertinent to the content featured in customer reviews. Nevertheless, not so much is known regarding the efficacy of personalized management responses in customer satisfaction. This study brings a new dimension to a previous study conducted by Zhang et al. (2020) where they propose topic-matched responses have a positive effect on future hotel ratings. In their study, a customer review can only contain one topic from predefined five topics. Then their model reveals corresponding management response's probability to include the same topic as customer review. However, customers generally voice out more than one aspect of a hotel where they accommodate so a customer review usually includes more than one topic in fact. By using the same review topics proposed by Zhang et al. (2020) in the phase of data annotation, this study presents a new perspective to personalized management responses by introducing multitopic matching responses. As such, with this perspective, a customer review and a management response can include more than one topic at the same time with the multilabel text classification approach. In addition, the empirical examination of this study reveals that topic matching degree has a positive effect on the growth of subsequent ratings, which is in line with findings in Zhang et al. (2020). This might suggest that higher topic matching degree indicates managers pay close attention to customer reviews, which is likely to enhance customer perceptions of hotels and lead to higher rating growth in subsequent periods.

As for the response ratio, the findings in this study are consistent with the literature (Xie et al., 2016; Sheng, 2018; Zhang et al., 2020). That is hotels with a higher response ratio is more likely to see positive changes on future ratings. Higher response ratio, as an indicator of a continual engagement in online platforms, might improve public recognition of a hotel's endeavour to interact with customers. With respect to response length, this study as consistent with Zhang et al. (2020) shows that longer responses have a positive effect on rating growth in monthly period. However, there is no significant effect of response length on rating growth found in weekly period. Besides, there is no adequate statistical evidence in this study showing that response promptness has a positive effect on rating growth in the monthly period which is consistent with Min et al. (2015) who report no effect of response speed on customer ratings, unlike Zhang et al. (2020), who find quicker response leads to a rating growth increase in the monthly period. Yet it is also shown in this study that response speed is influential in rating growth in a weekly period. In terms of usefulness of response speed, this is coherent with Sheng (2018) who report that faster responses are positively influencing average customer ratings in weekly duration. Consequently, the length of response text is effective in a monthly period while the importance of response speed is recognized in a weekly period. It is also worth to rearticulate that the positive influence of both topic matching degree and response length in rating growth are recognized in this study. In this case, potential customers are not only affected by the length of the response text, but they also place a substantial value on how consistent the content of response is with the reviews.

In addition to the above findings which consists of response attributes, this study presents the moderating effects of valence, variance, and volume of online reviews on the efficacy of topic-matched responses. As in line with Zhang et al. (2020), the effect of topic-matched responses on rating growth for hotels with high review valence, and variance is weaker when compared to the hotels with low review valence and variance. These results can be elucidated as positive information often presents greater anticipation and diagnosis, which can cause customers to take heed of shortcomings in service more and reduce the efficacy of topic-matched response. Additionally, high variability of reviews can cause customers to face more unfavourable reviews on the platform. This can also weaken the influence of responses on satisfaction. The volume of reviews is also negatively moderating the influence of topic-matched responses. This can be interpreted as the abundance of online reviews can lead customers to pay more attention to reviews to find sufficient information rather than management responses, which can also negatively impact the influence of management responses on hotel rating increase.

This research adds to the existing literature on the knowledge-based view of a firm (Grant, 1996; Nonaka, 1994) by exploring how businesses use social media data to innovate new ways of reacting to customers. This study also enhances dynamic marketing capabilities. It sheds light on how to increase future ratings through the more strategic use of response styles. It emphasizes how businesses can improve customer relationship management by showing the efficacy of personalized responses in increasing subsequent ratings.

CHAPTER 6:
CONCLUSION

6.1 Introduction

This section concludes the thesis with an overview of the research. Accordingly, the following part revisits the research aim and objectives with the discussion of the achievement of each objective. Then the next part summarizes the main research findings. The following part introduces the theoretical and practical implications of the research. Finally, research limitations and future research directions are presented in the last part of the chapter.

6.2 Meeting the Research Aim and Objectives

The aim of the research, as stated in Chapter 1, is to investigate strategic managerial responses to online reviews. To attain this aim, five research objectives are presented as follows.

6.2.1 Objective 1

The first objective is to review literature based on management responses to reviews, and SMA with its applications. Within this context, WOM and eWOM were initially reviewed, and differences between those were presented. Thereafter, an overview of social media and its relationship with CRM was mentioned. Management responses to online reviews was then examined along with exploring previous studies, and the research gap was identified. In the second part of the literature review, SMA techniques including SNA, sentiment analysis, and mostly text mining and NLP were presented. After that, advanced analytics consisting of machine learning and deep learning were reviewed. Following, two relevant theories were presented: Dynamic Capabilities and Knowledge-Based Theory.

6.2.2 Objective 2

The second objective was to reveal response strategies to online negative reviews based on response content using social media big data analytics. To define response strategies, management and linguistic studies were firstly reviewed, and then a sample from collected data was manually annotated. Consequently, five categories were identified: brand positioning, admittance inclusive, action inclusive, explanatory, and direct contact inclusive responses.

After data pre-processing, a thousand responses for each defined response content were annotated to be used as training sets in text classification. With the binary text classification approach, two machine learning and two deep learning methods were applied separately for each response content. Which method to apply for each response content was determined according to the best classification performances, which was presented in Chapter 4. After classifying all responses, a comprehensive exploratory analysis was conducted to reveal correlation between response content and overall hotel rating and differences between response strategies of 4-star and 5-star hotels.

6.2.3 Objective 3

The third objective was to identify effective response strategies on the ratings of returning customers.

The second phase of Chapter 4 presented the impact of managerial responses to subsequent ratings of returning customers who revisited the same hotel after a negative experience. Unlike the first study, the action inclusive category was divided into two as generic and specific action inclusive responses. Additionally, brand positioning category was not examined since the aim was to investigate managerial perspectives on customers’ specific problems. A logistic regression analysis was conducted to reveal which response strategies are effective in increasing customers’ subsequent rating.

6.2.4 Objective 4

The fourth objective was to investigate the effect of personalized responses on rating growth, revealing topic matching degree between responses and reviews (including both positive and negative reviews).

Thus, multitopic matching degrees between responses and corresponding reviews were calculated after classifying responses and reviews into five common topics with a deep learning assisted multilabel text classification approach. As presented in Chapter 5, an empirical model was created for the purpose of analysing whether multitopic matched responses are influential in increasing subsequent ratings.

6.3 Summary of Results

The research questions whose answer was sought in the first phase of Study–1 is as follows: What strategies managers follow when responding to online negative reviews?

The first phase of this study was conducted adopting social media big data analytics with the adoption of various text classification methods. On a large data set, it was investigated whether the response content was correlated with the overall hotel rating, the differences in the response strategies of 4-star and 5-star hotels, and whether the quantitative attributes of the responses were correlated with the overall hotel rating. Based on these research question, the main results of the first phase of Study–1 are as follows:

- The correlation analysis between response content and overall hotel ratings indicates that admittance of mistake, brand positioning, and direct contact request inclusive response strategies have significant and positive relationship with overall hotel rating, whereas there is no significant relationship found between action inclusive and explanatory response strategies and overall hotel rating.
- The comparative analysis of 4-star and 5-star hotels for the usage of five response categories shows that the use of these response strategies is mostly greater in the responses of 5-star hotels. However, explanatory response is the only response strategy that is seen more in the responses of 4-star hotels than 5-star hotels.
- The average response rate of 5-star hotels is higher than the average response of 4-star hotels. The median of response speed of 4-star hotels is higher than the median of response speed of 5-star hotels. The median response length of 5-star hotels is greater than 4-star hotels.
- There is no correlation found between the response speed and overall hotel rating.
- There is a positive relationship between response rate to negative reviews and overall hotel rating.

- There is a positive relationship between response length to negative reviews and overall hotel rating. This correlation is the strongest compared to correlations of response ratio and response speed with overall hotel rating.

The second question of this research was: What are the effective response strategies increasing subsequent rating of returning customers who post negative rating in their previous visit?

In the second phase of Study–1, management responses were manually read and coded based on generic action inclusive, specific action inclusive, direct contact request inclusive, admittance inclusive, and explanatory categories. According to the research question, the main results of Study–2 are as follows:

- Admittance inclusive response has a positive and significant relationship with the increase of subsequent rating of returning customers. Provision of admittance inclusive response is almost six times more likely to increase the subsequent rating compared to the response lacking admittance statement.
- Specific action inclusive response has a positive and significant relationship with the increase of subsequent rating of returning customers. Responses including specific action statement are five times more likely to increase the subsequent rating of returning customers compared to the response that is lack of specific action content.
- Direct contact inclusive response has a positive and significant relationship with the increase of subsequent rating of returning customers. Provision of direct contact inclusive response is five times more likely to increase the next rating of returning customers compared to the response not including direct contact request statement.
- General action inclusive and explanatory responses do not have significant impact on subsequent ratings of returning customers.
- There is no significant relationship found between the length and speed of the response and increase in subsequent rating of returning customers.

The third question of this research was: Do personalized responses with multitopic matching degree have effect on subsequent ratings?

Study–2 in this research examined the influence of multitopic matching degree between responses and reviews on rating growth of hotels. A summary of findings of Study–2 follows:

- Multitopic matching degree between response and reviews has positive effect on subsequent ratings of customers.
- Higher response rate leads the higher rating growth in the subsequent period.
- There is a positive and significant moderating effect of hotel star on topic matching degree, suggesting the influence of topic matching of management response in rating growth is stronger in higher grade hotels.
- The moderating effects of all review characteristics (valence, variance, and volume) decrease the influence of topic matching on rating growth.
- Response length is positively influential in rating growth in monthly period while response speed is effective in growing online ratings in weekly period.

6.4 Theoretical Contributions

This doctoral thesis is motivated with the keen on proving how social media is a valuable source of big data where businesses can strategize user generated content via engagement in online networking sites. To find the answers to research questions, two studies are conducted to investigate management responses in online social interactions by presenting novel perspectives. Overall, this thesis has theoretical contributions mentioned below.

This research expands the current research of knowledge-based view of a firm, which claims that knowledge is a key resource for firms striving to develop new sources of competitive advantage. The presence of a data-rich environment necessitates firms to leverage various data sources to develop new knowledge and therefore optimize operations (Khan & Vorley, 2017; Tian, 2017). Capturing knowledge via text analytics of online reviews and delivering responses are crucial for the highly dynamic and competitive hotel industry to improve customer satisfaction and hotel performance (Liu et al., 2018; Xie et al., 2014). This research highlights how firms could leverage the potential knowledge generated through online reviews and responses and enhance service quality and customers' satisfaction. Rather than deploying traditional approaches such as survey or interviews, this research adopts big data analytics through social media, natural language processing, and advanced deep learning techniques with a mixed-method design. This research not only presents the efficacies of the deep learning and natural language processing technology in generating knowledge and developing pertinent strategies of managerial responses, but also demonstrate the usefulness of data-driven strategies for businesses in extremely competitive business environment.

Findings from this thesis support the concept of the dynamic marketing capabilities model (Barrales-Molina et al., 2014). It is demonstrated that obtaining marketing information via data mining and analysis of consumer generated information underpins businesses to grasp a better understanding of what is desired in the market and take appropriate actions on service improvement, thereby improving performance and enhancing customer satisfaction. Developing an idea, product, or service to meet future demands rapidly and efficiently alongside responding to customer needs enables firms to stay ahead of competitors in a dynamic market (Day, 2011; Xu et al., 2018). This research highlights effective communication styles when interacting with customers for the improvement of strategic decision-making and customer relationship management. This also supports the notion of dynamic marketing capability by putting emphasis on the combination of social media and digital marketing strategies to enhance relationship management capabilities between businesses and customers (Wang & Kim, 2017).

Study-1 in this research adds weight to the literature on management's responses to negative reviews, with a particular focus on exploration of strategic response content. Unlike current research, which mostly focuses on a simple classification of responses (accommodativeness vs. defensiveness, generic vs. specific), this study presents a detailed analysis of what content strategies hotels adopt when responding to dissatisfied customers. With the help of natural language processing and deep learning techniques, a comprehensive classification of management response content to negative reviews is conducted based on five strategic moves (action, admittance, brand positioning, direct contact request, and explanatory). Visually assisted comparative and descriptive analyses enhance our understanding of managerial reactions in case of a service failure and reveal that the response strategies to negative reviews differ according to the hotel class and average rating. Moreover, the effect of management responses to subsequent ratings of returning customers has not been widely examined in the literature. Although existing research demonstrate that the response provision has a positive effect on future ratings of repeated customers (Gu & Ye, 2014), what response content strategies can be influential in increasing ratings of dissatisfied returning customers is still open to investigate. The second phase of the Study–1 reveals effective response strategies to increase subsequent rating of dissatisfied customers. Overall, the Study–1 contributes to management responses to negative review literature with the illustration of useful insights for improvement of customer satisfaction and brand reputation.

Additionally, Study-2 acknowledges that big data from social media can help businesses deliver personalized customer relationship for each customer. Even though studies show the importance of customizing management response content based on the review (Li et al., 2018; Wang & Chaudhry, 2018), there is limited research presenting the effect of personalized management responses in future ratings. Zhang et al. (2020) recently demonstrate that personalized responses can increase subsequent ratings of customers by controlling topic matching degree between responses and reviews. However, customers frequently voice their sentiments about more than one aspect of service they experience. The second study extends their study by adopting the multitopic matching degree approach between responses and reviews and removes this barrier. Unlike their approach, the approach in the second study does not restrict topic matching degree calculation based on one matched topic between a response and review. Instead, the number of matched topics between a response and the review is calculated. In addition, it is found that higher multitopic matching degree leads to higher overall rating. This can suggest the strategic value of management responses and the importance of personalized customer relationship management in customer satisfaction. Finally, compared to Study–1 in this thesis and several studies that focus on responses to only negative reviews, Study–2 in this thesis does not limit management response strategies to negative reviews but also investigates their effects on positive reviews.

6.5 Practical Implications

From a practical perspective, this research provides marketers valuable insights in terms of engagement strategies in online social interactions. Active engagement in the interactive online communications can offer diverse benefits for businesses. Interactions between business and customers on the internet are open to public and actions that taken by businesses can be more illustrative than customer to customer interactions for potential customers. In this sense, businesses are better to develop strategies to enhance their visibility and effective communication styles.

A direct form of firm–customer interactions is managerial responses to online reviews. Management response is a useful tool for firms to transfer information, provide customer support, and handle customer comments. Considering the potential detrimental impacts of negative reviews on business performance and brand image, managers are suggested to develop proper strategies to mitigate these impacts and learn from customer complaints to improve service. The first study in this research suggests managers to strategize their responses to negative reviews in terms of content, speed, and the text length. Although this study does not demonstrate the direct effect of this strategies to negative reviews in overall ratings, it might be beneficial to keep in mind for firms that hotels with the higher customer ratings have a higher response ratio and mostly respond by admitting the responsibility if they are mistaken, positioning their brand, and requesting direct contact with the customers.

The second phase of the first study in this research presents some strategies for managers when responding to returning customers with negative experiences in their previous visit. Using generic action statement does not have significant effect, whereas indicating that specific action has been taken or will be taken regarding the negative experience of the customer is significantly more effective in increasing subsequent rating of returning customer. Hence, it is recommended that managers make a specific action statement related to the corresponding customer's negative experience in their response. In addition, responses that acknowledge the firm's fault are much more effective in increasing the next rating of returning customers than responses that do not. This implies that if the firm is mistaken, it is recommended to use a statement that accepts the responsibility for the fault.

Last, the request to contact the customer directly is effective in increasing the subsequent rating of that customer. This might imply that direct contact with customers after replying to their reviews might improve customer perception of importance given to customer relations.

In addition, findings in the second study demonstrate the positive influence of personalized responses on future ratings. In this regard, managers should seriously comprehend what topics are involved in the relevant review that is regardless of positive and negative and attach a response that consists of as many matched topics as the corresponding review does rather than providing generic or standardized responses. Keeping this personalized engagement and communication strategy as a consistent practice could enhance rating growth on online social networking sites, which can implicitly attract purchase decisions of potential customers and enhance brand prestige. Apart from the response content, it is highly recommended that response providers take care of responding on a regular basis, response promptness, and response length that sufficiently long to include the customer's comments and concerns to receive favourable outcomes in the future ratings. The multitopic matching degree approach can be adopted to be used in an intelligence assisted application by hotel managers to automatically reveal what topics involved in customer reviews. Hence, response provider can tailor the content of a response based on what topics a corresponding review includes. This might effectively help firms to improve response quality and gain considerable amount of time.

Finally, findings from this research supports that social media big data analytics and deep learning technologies have potency to acquire actionable insights and enhance the decision-making process with the utilization of knowledge from market. Adopting data-driven business strategies help organizations keep up with the ever-changing marketing environment. Investing in information technology and data talent is crucial to gain and sustain big data competencies. Having state-of-the-art IT systems for data analytics, equipping organizations with data analysts, business analysts, and social media marketing specialists can lead to developing strategic plans and find solutions to specific problems through ongoing knowledge management.

6.6 Limitations and Future Directions

Despite the theoretical and practical contributions this research makes, the following limitations can be addressed in and offer possible directions for future research.

As for the first phase of Study–1, management responses to negative reviews are investigated only for 4-star and 5-star hotels in three popular cities. So as to make a more extensive comparison about response strategies, it would be more informative to capture response data from lower graded hotels and more cities. Importantly, only management responses are investigated without controlling the content of negative reviews. Future research can examine these response strategies by categorizing online negative reviews based on complaint reasons and find out which response strategies are more effective on which specific complaint. In addition, although this study presents a set of highlights such as response strategy differences of hotels with different grades and correlational analysis of response strategies with overall rating, it does not provide the influence of responses on future ratings. Future research could analyse the impact of these response strategies to negative reviews on subsequent ratings.

The second phase of the Study–1 investigates effective response strategies to increase returning customers' ratings. However, only responses and reviews of 5-star hotels in London are explored. Future research can investigate more hotels with different grades from more cities and compare the effect of response strategies based on hotel grades. In addition, the model used in this study includes only response characteristics such as content, speed, and length. However, there might be other observable (review characteristics such as volume, variety, and valence as well as the career position of the response provider) or unobservable (e.g., the customer may be given a special promotion for a future visit that is not mentioned on the response or might have a membership card) factors leading to an increase in the next rating of previously dissatisfied customers. These factors can be added in future research, and their effect on the next rating of returning customers could be controlled.

The second study in this research has several limitations that can be addressed in future research. In the empirical model, which mainly aims to reveal the effect of multitopic matching degree between response and reviews on future ratings, if available, sales data can be added to examine the impact of multitopic matching degree on business performance. There are also some factors that might possibly affect the influence of management responses that are not included in the empirical model. These include the career position and work experience of response provider, internal strategy of firm in social media management, price information of service, and demographic information of customers. These factors could be added in the future research when related data is available. Plus, our data are collected from 3, 4, and 5-star London hotels from a single review website. Findings from one city and one travelling website may not be generalizable to other hotel market, the availability of larger data from various locations involving 1 and 2-star hotels would provide a more extensive viewpoint of the interrelationships between online reviews, hotel ratings, and management responses.

Last, findings from the two studies can be further affirmed and generalized with new data and alternative methods. This doctoral research, for example, analyses managerial responses and customer reviews from the tourism and hospitality sector. This research setting can be extended other social networking platforms such as Facebook and Twitter, where a faster information flow takes place. It might also lead to fruitful research to analyse organizational reactions to negative comments and assess the effect of firm engagement using data from other sectors such as education, retailing, health care, and sport. This would be intriguing to compare the results and differences of firm engagement in organizational level from different sectors.

REFERENCES

Affendey, S., & Mamat, A. (2015). A systematic review on the profiling of digital news portal for Big Data veracity. *Procedia computer science*, *72*, 390-397.

Aggarwal, C. C. (2015). Mining text data. Data mining,

Ajzen, I., Fishbein, M., Lohmann, S., & Albarracín, D. (2018). The influence of attitudes on behavior. *The handbook of attitudes*, 197-255.

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International journal of production economics*, *182*, 113-131. https://doi.org/10.1016/j.ijpe.2016.08.018

Alalwan, A. A., Rana, N. P., Dwivedi, Y. K., & Algharabat, R. (2017). Social media in marketing: A review and analysis of the existing literature. *Telematics and informatics*, *34*(7), 1177-1190. https://doi.org/10.1016/j.tele.2017.05.008

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET),

Alexander, B. (2006). Web 2.0: A new wave of innovation for teaching and learning? *Educause review*, *41*(2), 33-34.

Alhabash, S., & Ma, M. (2017). A Tale of Four Platforms: Motivations and Uses of Facebook, Twitter, Instagram, and Snapchat Among College Students? *Social media + society*, *3*(1), 205630511769154. https://doi.org/10.1177/2056305117691544

Alharahsheh, H. H., & Pius, A. (2020). A review of key paradigms: Positivism VS interpretivism. *Global Academic Journal of Humanities and Social Sciences*, *2*(3), 39-43.

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). Text Summarization Techniques: A Brief Survey.

Amplayo, R. K., Bražinskas, A., Suhara, Y., Wang, X., & Liu, B. (2022). Beyond Opinion Mining: Summarizing Opinions of Customer Reviews.

Anderson, E. W. (1998). Customer satisfaction and word of mouth. *Journal of Service Research*, *1*(1), 5-17.

Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2019). The future of social media in marketing. *Journal of the Academy of Marketing Science*, *48*(1), 79-95. https://doi.org/10.1007/s11747-019-00695-1

Argote, L., & Ingram, P. (2000). Knowledge Transfer: A Basis for Competitive Advantage in Firms. *Organizational behavior and human decision processes*, *82*(1), 150-169. https://doi.org/10.1006/obhd.2000.2893

Arndt, J. (1967). Word-of-mouth advertising and informal communication. *Risk taking and information handling in consumer behavior*, 188-239.

Asgari-Chenaghlu, M., Feizi-Derakhshi, M. R., Farzinvash, L., Balafar, M. A., & Motamed, C. (2021). CWI: A multimodal deep learning approach for named entity recognition from social media using character, word and image features. *Neural computing & applications*, *34*(3), 1905-1922. https://doi.org/10.1007/s00521-021-06488-4

Atuahene-Gima, K. (1995). An exploratory analysis of the impact of market orientation on new product performance a contingency approach. *The Journal of product innovation management*, *12*(4), 275-293. https://doi.org/10.1016/0737-6782(95)00027-Q

Ayeh, J. K., Au, N., & Law, R. (2013). "Do We Believe in TripAdvisor?" Examining Credibility Perceptions and Online Travelers' Attitude toward Using User-Generated Content. *Journal of travel research*, *52*(4), 437-452. https://doi.org/10.1177/0047287512475217

Azarbonyad, H., & Marx, M. (2019). How Many Labels? Determining the Number of Labels in Multi-Label Text Classification. In (pp. 156-163). Cham: Cham: Springer International Publishing.

Azemi, Y., Ozuem, W., & Howell, K. E. (2020). The effects of online negative word-of-mouth on dissatisfied customers: A frustration–aggression perspective. *Psychology & marketing*, *37*(4), 564-577. https://doi.org/10.1002/mar.21326

Babic Rosario, A., Sotgiu, F., de Valck, K., & Bijmolt, T. H. A. (2016). The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors. *Journal of marketing research*, *53*(3), 297-318. https://doi.org/10.1509/jmr.14.0380

Bai, L., & Yan, X. (2020). Impact of firm-generated content on firm performance and consumer engagement: Evidence from social media in China. *Journal of electronic commerce research*, *21*(1), 56-74.

Baldi, P. (2012). Autoencoders, unsupervised learning, and deep architectures. Proceedings of ICML workshop on unsupervised and transfer learning,

Ballings, M., & Van den Poel, D. (2015). CRM in social media: Predicting increases in Facebook usage frequency. *European journal of operational research*, *244*(1), 248-260. https://doi.org/10.1016/j.ejor.2015.01.001

Bambauer-Sachse, S., & Mangold, S. (2011). Brand equity dilution through negative online word-of-mouth communication. *Journal of retailing and consumer services*, *18*(1), 38-45. https://doi.org/10.1016/j.jretconser.2010.09.003

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management*, *17*(1), 99-120.

Barney, J. B., Ketchen, D. J., & Wright, M. (2011). The Future of Resource-Based Theory: Revitalization or Decline? *Journal of management*, *37*(5), 1299-1315. https://doi.org/10.1177/0149206310391805

Barrales-Molina, V., Martínez-López, F. J., & Gázquez-Abad, J. C. (2014). Dynamic Marketing Capabilities: Toward an Integrative Framework: Dynamic Marketing Capabilities. *International journal of management reviews : IJMR*, *16*(4), 397-416. https://doi.org/10.1111/ijmr.12026

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information fusion*, *28*, 45-59. https://doi.org/10.1016/j.inffus.2015.08.005

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation Learning: A Review and New Perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798-1828. https://doi.org/10.1109/TPAMI.2013.50

Berger, L. A. (1988). Word-of-mouth reputations in auto insurance markets. *Journal of Economic Behavior and Organization*, *10*(2), 225-234. https://doi.org/10.1016/0167-2681(88)90046-7

Blom, J. o., & Monk, A. F. (2003). Theory of Personalization of Appearance: Why Users Personalize Their PCs and Mobile Phones. *Human-computer interaction*, *18*(3), 193-228. https://doi.org/10.1207/S15327051HCI1803_1

Bogner, W. C., & Bansal, P. (2007). Knowledge Management as the Basis of Sustained High Performance. *Journal of management studies*, *44*(1), 165-188. https://doi.org/10.1111/j.1467-6486.2007.00667.x

Bohlouli, M., Dalter, J., Dornhöfer, M., Zenkert, J., & Fathi, M. (2015). Knowledge discovery from social media using big data-provided sentiment analysis (SoMABiT). *Journal of information science*, *41*(6), 779-798.

Borek, A., Parlikad, A. K., Webb, J., & Woodall, P. (2013). *Total information risk management: maximizing the value of data and information assets*. Newnes.

Borges, M., Hoppen, N., & Luce, F. B. (2009). Information technology impact on market orientation in e-business. *Journal of Business Research*, *62*(9), 883-890. https://doi.org/10.1016/j.jbusres.2008.10.010

Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research*, *70*, 328-337. https://doi.org/10.1016/j.jbusres.2016.08.006

Brodie, R. J., Ilic, A., Juric, B., & Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research*, *66*(1), 105-114. https://doi.org/10.1016/j.jbusres.2011.07.029

Brownlee, J. (2020). *Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python*. Machine Learning Mastery.

Bryman, A., & Bell, E. (2015). *Business research methods* (Fourth edition. ed.). Oxford University Press.

Bukovina, J. (2016). Social media big data and capital markets—An overview. *Journal of behavioral and experimental finance*, *11*, 18-26. https://doi.org/10.1016/j.jbef.2016.06.002

Burrell, G., & Morgan, G. (2017). *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life*. Routledge.

Buttle, F. A. (1998). Word of mouth: understanding and managing referral marketing. *Journal of strategic marketing*, *6*(3), 241-254. https://doi.org/10.1080/096525498346658

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. Proceedings of the 23rd international conference on Machine learning,

Cavdar Aksoy, N., Tumer Kabadayi, E., Yilmaz, C., & Kocak Alan, A. (2021). A typology of personalisation practices in marketing in the digital age. *Journal of marketing management*, *37*(11-12), 1091-1122. https://doi.org/10.1080/0267257X.2020.1866647

Chang, H. H., & Wu, L. H. (2014). An examination of negative e-WOM adoption: Brand commitment as a moderator. *Decision Support Systems*, *59*(1), 206-218. https://doi.org/10.1016/j.dss.2013.11.008

Chang, W., Park, J. E., & Chaiy, S. (2010). How does CRM technology transform into organizational performance? A mediating role of marketing capability. *Journal of Business Research*, *63*(8), 849-855. https://doi.org/10.1016/j.jbusres.2009.07.003

Chen, C.-H., Nguyen, B., Klaus, P. P., & Wu, M.-S. (2015). Exploring Electronic Word-of-Mouth (eWOM) in The Consumer Purchase Decision-Making Process: The Case of Online Holidays - Evidence from United Kingdom (UK) Consumers. *Journal of travel & tourism marketing*, *32*(8), 953-970. https://doi.org/10.1080/10548408.2014.956165

Chen, C. P., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, *275*, 314-347.

Chen, Z., & Lurie, N. H. (2013). Temporal contiguity and negativity bias in the impact of online word of mouth. *Journal of marketing research*, *50*(4), 463-476. https://doi.org/10.1509/jmr.12.0063

Cheung, C. M., & Thadani, D. R. (2010). The Effectiveness of Electronic Word-of-Mouth Communication: A Literature Analysis. *Bled eConference*, *23*, 329-345.

Cheung, C. M. K., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems*, *54*(1), 461-470. https://doi.org/10.1016/j.dss.2012.06.008

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of marketing research*, *43*(3), 345-354. https://doi.org/10.1509/jmkr.43.3.345

Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing science (Providence, R.I.)*, *29*(5), 944-957. https://doi.org/10.1287/mksc.1100.0572 (Marketing Science)
Choi, B., & Choi, B.-J. (2014). The effects of perceived service recovery justice on customer affection, loyalty, and word-of-mouth. *European journal of marketing*, *48*(1/2), 108-131. https://doi.org/10.1108/EJM-06-2011-0299
Choi, T.-M., Chan, H. K., & Yue, X. (2017). Recent Development in Big Data Analytics for Business Operations and Risk Management. *IEEE transactions on cybernetics*, *47*(1), 81-92. https://doi.org/10.1109/TCYB.2015.2507599
Choudhury, M. M., & Harrigan, P. (2014). CRM to social CRM: the integration of new technologies into customer relationship management. *Journal of strategic marketing*, *22*(2), 149-176. https://doi.org/10.1080/0965254X.2013.876069
Chowdhary, K. (2020). Natural language processing. In *Fundamentals of Artificial Intelligence* (pp. 603-649). Springer.
Collis, J., & Hussey, R. (2003). *Business research : a practical guide for undergraduate and postgraduate students* (2nd ed.). Palgrave Macmillan.
Collis, J., & Hussey, R. (2013). *Business research: A practical guide for undergraduate and postgraduate students*. Macmillan International Higher Education.
Coltman, T. (2007). Why build a customer relationship management capability? *The journal of strategic information systems*, *16*(3), 301-320. https://doi.org/10.1016/j.jsis.2007.05.001
Commission, F. B. D. (2012). *Demystifying big data: A practical guide to transforming the business of government*.
Constantinides, E. (2014). Foundations of Social Media Marketing. *Procedia, social and behavioral sciences*, *148*, 40-57. https://doi.org/10.1016/j.sbspro.2014.07.016
Constantiou, I. D., & Kallinikos, J. (2015). New Games, New Rules: Big Data and the Changing Context of Strategy. *Journal of Information Technology*, *30*(1), 44-57. https://doi.org/10.1057/jit.2014.17
Creswell, J. W., & Creswell, J. D. (2018). *Research design : qualitative, quantitative, and mixed methods approaches* (Fifth edition. ed.). SAGE Publications, Inc.
Cui, G., Lui, H.-K., & Guo, X. (2012). The Effect of Online Consumer Reviews on New Product Sales. *International Journal Of Electronic Commerce*, *17*(1), 39-58. https://doi.org/10.2753/JEC1086-4415170102
Day, G. S. (2011). Closing the Marketing Capabilities Gap. *Journal of Marketing*, *75*(4), 183-195. https://doi.org/10.1509/jmkg.75.4.183
De Cock, C., & Land, C. (2006). Organization/Literature: Exploring the Seam. *Organization studies*, *27*(4), 517-535. https://doi.org/10.1177/0170840605058234
Dellarocas, C. (2003). The Digitization of Word-of-mouth: Promise and Challenges of Online Feedback Mechanisms. *IDEAS Working Paper Series from RePEc*.
Dellarocas, C., & Narayan, R. (2007). Tall heads vs. long tails: Do consumer reviews increase the informational inequality between hit and niche products? *Robert H. Smith School of Business Research Paper*(06-056).
Deng, L., & Liu, Y. (2018). *Deep learning in natural language processing*. Springer.
Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Diebold, F. X., Cheng, X., Diebold, S., Foster, D., Halperin, M., Lohr, S., Mashey, J., Nickolas, T., Pai, M., & Pospiech, M. (2012). A Personal Perspective on the Origin (s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline∗.
Dudovskiy, J. (2016). The ultimate guide to writing a dissertation in business studies: A step-by-step assistance. *Pittsburgh, USA*, *51*.
Duong, C. T. P. (2020). Social Media. A Literature Review. *Journal of Media Research*, *13*(3).
East, R., Hammond, K., & Lomax, W. (2008). Measuring the impact of positive and negative word of mouth on brand purchase probability. *International journal of research in marketing*, *25*(3), 215-224. https://doi.org/10.1016/j.ijresmar.2008.04.001
Edgar, T. W., & Manz, D. O. (2017). *Research methods for cyber security*. Syngress.
Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: what are they? *Strategic management journal*, *21*(10-11), 1105-1121.
El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In *Machine Learning in Radiation Oncology* (pp. 3-11). Springer.
Elaraby, N. M., Elmogy, M., & Barakat, S. (2016). Deep learning: Effective tool for big data analytics. *International Journal of Computer Science Engineering (IJCSE)*, *9*.
Elhoseny, M., Kabir Hassan, M., & Kumar Singh, A. (2020). Special issue on cognitive big data analytics for business intelligence applications: Towards performance improvement. *International journal of information management*, *50*, 413-415. https://doi.org/10.1016/j.ijinfomgt.2019.08.004
Ellison, N. B., & Boyd, D. (2013). Sociality through social network sites. *The Oxford handbook of internet studies*, *151*, 172.
Elwalda, A., Lü, K., & Ali, M. (2016). Perceived derived attributes of online customer reviews. *Computers in Human Behavior*, *56*, 306-319. https://doi.org/10.1016/j.chb.2015.11.051
Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897-904. https://doi.org/10.1016/j.jbusres.2015.07.001
Fan, W., & Gordon, M. (2014). The power of social media analytics. *Communications of the ACM*, *57*(6), 74-81. https://doi.org/10.1145/2602574
Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism management (1982)*, *52*, 498-506. https://doi.org/10.1016/j.tourman.2015.07.018
Farsi, D. (2021). Social media and health care, part i: Literature review of social media use by health care providers. *Journal Of Medical Internet Research*, *23*(4), e23205-e23205. https://doi.org/10.2196/23205
Ferguson, P., O'Hare, N., Davy, M., Bermingham, A., Sheridan, P., Gurrin, C., & Smeaton, A. F. (2009). Exploring the use of paragraph-level annotations for sentiment analysis of financial blogs.
Filo, K., Lock, D., & Karg, A. (2015). Sport and social media research: A review. *Sport management review*, *18*(2), 166-181. https://doi.org/10.1016/j.smr.2014.11.001
Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., & Freling, T. (2014). How Online Product Reviews Affect Retail Sales: A Meta-analysis. *Journal of retailing*, *90*(2), 217-232. https://doi.org/10.1016/j.jretai.2014.04.004
Fortunato, S. (2010). Community detection in graphs. *Physics reports*, *486*(3-5), 75-174. https://doi.org/10.1016/j.physrep.2009.11.002
Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study.

International journal of production economics, *165*, 234-246. https://doi.org/10.1016/j.ijpe.2014.12.031
Fowler, J. H., & Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *BMJ*, *337*(dec04 2), a2338-a2338. https://doi.org/10.1136/bmj.a2338
Franks, H., Griffiths, N., & Anand, S. S. (2014). Learning agent influence in MAS with complex social networks. *Autonomous agents and multi-agent systems*, *28*(5), 836-866. https://doi.org/10.1007/s10458-013-9241-1
Gabriel, Y., Gray, D. E., & Goregaokar, H. (2013). Job loss and its aftermath among managers and professionals: wounded, fragmented and flexible. *Work, employment and society*, *27*(1), 56-72. https://doi.org/10.1177/0950017012460326
Gamallo, P., & Garcia, M. (2019). Editorial for the Special Issue on “Natural Language Processing and Text Mining”. In: Multidisciplinary Digital Publishing Institute.
Gambhir, M., & Gupta, V. (2016). Recent automatic text summarization techniques: a survey. *The Artificial intelligence review*, *47*(1), 1-66. https://doi.org/10.1007/s10462-016-9475-9
Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137-144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007
Garaus, M., Wagner, U., & Kummer, C. (2015). Cognitive fit, retail shopper confusion, and shopping value: Empirical investigation. *Journal of Business Research*, *68*(5), 1003-1011. https://doi.org/10.1016/j.jbusres.2014.10.002
Gargiulo, F., Silvestri, S., Ciampi, M., & De Pietro, G. (2019). Deep neural network for hierarchical extreme multi-label text classification. *Applied soft computing*, *79*, 125-138. https://doi.org/10.1016/j.asoc.2019.03.041
Gasparetti, F., Micarelli, A., & Sansonetti, G. (2018). Community Detection and Recommender Systems. In.
Gaur, A. S., Kumar, V., & Singh, D. (2014). Institutions, resources, and internationalization of emerging economy firms. *Journal of world business : JWB*, *49*(1), 12-20. https://doi.org/10.1016/j.jwb.2013.04.002
Geerts, W. (2018). *Top 100 City Destinations 2018*. E. International. https://go.euromonitor.com/white-paper-travel-2018-100-cities.html?utm_source=blog%26utm_medium=blog%26utm_campaign=CT_WP_18_12_04_100%20Cities%26utm_content=organic
George, G., Haas, M., & Pentland, A. (2014). Big Data and Management. *Academy of Management journal*, *57*(2), 321-326. https://doi.org/10.5465/amj.2014.4002
Gerard, G., Martine, R. H., & Alex, P. (2014). FROM THE EDITORS: BIG DATA AND MANAGEMENT. *Academy of Management journal*, *57*(2), 321-326.
Ghani, N. A., Hamid, S., Targio Hashem, I. A., & Ahmed, E. (2019). Social media big data analytics: A survey. *Computers in Human Behavior*, *101*, 417-428. https://doi.org/10.1016/j.chb.2018.08.039
Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., & Smith, N. A. (2010). *Part-of-speech tagging for twitter: Annotation, features, and experiments*.
Goldsmith, R. E., & Horowitz, D. (2006). Measuring motivations for online opinion seeking. *Journal of interactive advertising*, *6*(2), 2-14.
Grant, R. M. (1996). Toward a Knowledge-Based Theory of the Firm. *Strategic management journal*, *17*(S2), 109-122. https://doi.org/10.1002/smj.4250171110

Greenberg, P. (2010). *CRM at the speed of light: Social CRM strategies, tools, and techniques*. McGraw-Hill New York.
Greenberg, P. (2010). The impact of CRM 2.0 on customer insight. *The Journal of business & industrial marketing*, *25*(6), 410-419. https://doi.org/10.1108/08858621011066008
Gu, B., & Ye, Q. (2014). First Step in Social Media: Measuring the Influence of Online Management Responses on Customer Satisfaction. *Production and operations management*, *23*(4), 570-582. https://doi.org/10.1111/poms.12043
Gudivada, V. N., Rao, D., & Raghavan, V. V. (2015). Big data driven natural language processing research and applications. In *Handbook of Statistics* (Vol. 33, pp. 203-238). Elsevier.
Guo, B., & Zhou, S. (2016). Understanding the impact of prior reviews on subsequent reviews: The role of rating volume, variance and reviewer characteristics. *Electronic Commerce Research and Applications*, *20*, 147-158. https://doi.org/10.1016/j.elerap.2016.10.007
Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism management (1982)*, *59*, 467-483. https://doi.org/10.1016/j.tourman.2016.09.009
Gutierrez-Gutierrez, L. J., Barrales-Molina, V., & Kaynak, H. (2018). The role of human resource-related quality management practices in new product development: A dynamic capability perspective. *International journal of operations & production management*, *38*(1), 43-66. https://doi.org/10.1108/IJOPM-07-2016-0387
Harrison-Walker, L. J. (2001). The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment As Potential Antecedents. *Journal of Service Research*, *4*(1), 60-75. https://doi.org/10.1177/109467050141006
Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information systems (Oxford)*, *47*, 98-115. https://doi.org/10.1016/j.is.2014.07.006
Hayat, M. K., Daud, A., Alshdadi, A. A., Banjar, A., Abbasi, R. A., Bao, Y., & Dawood, H. (2019). Towards Deep Learning Prospects: Insights for Social Media Analytics. *IEEE access*, *7*, 36958-36979. https://doi.org/10.1109/ACCESS.2019.2905101
He, W., Wang, F.-K., & Akula, V. (2017). Managing extracted knowledge from big social media data for business decision making. *Journal of knowledge management*, *21*(2), 275-294. https://doi.org/10.1108/JKM-07-2015-0296
Heidemann, J., Klier, M., & Probst, F. (2012). Online social networks: A survey of a global phenomenon. *Computer networks (Amsterdam, Netherlands : 1999)*, *56*(18), 3866-3878. https://doi.org/10.1016/j.comnet.2012.08.009
Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, *18*(1), 38-52. https://doi.org/10.1002/dir.10073
Hennig-Thurau, T., Malthouse, E. C., Friege, C., Gensler, S., Lobschat, L., Rangaswamy, A., & Skiera, B. (2010). The impact of new media on customer relationships. *Journal of service research : JSR*, *13*(3), 311-330. https://doi.org/10.1177/1094670510375460
Hennig-Thurau, T., & Walsh, G. (2003). Electronic word-of-mouth: Motives for and consequences of reading customer articulations on the Internet. *International Journal Of Electronic Commerce*, *8*(2), 51-74.
Heron, J. (1996). *Co-operative inquiry: Research into the human condition*. Sage.
Hill, C. W., Jones, G. R., & Schilling, M. A. (2014). *Strategic management: Theory & cases: An integrated approach*. Cengage Learning.

Hinton, G. E. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science (American Association for the Advancement of Science)*, *313*(5786), 504-507. https://doi.org/10.1126/science.1127647

Hinton, G. E. (2009). Deep belief networks. *Scholarpedia*, *4*(5), 5947.

Ho, V. (2018). Exploring the effectiveness of hotel management's responses to negative online comments. *Lingua*, *216*, 47-63. https://doi.org/10.1016/j.lingua.2018.10.004

Hobbs, J. R., & Riloff, E. (2010). Information Extraction. *Handbook of natural language processing*, *15*, 16.

Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014). Consumer Brand Engagement in Social Media: Conceptualization, Scale Development and Validation. *Journal of Interactive Marketing*, *28*(2), 149-165. https://doi.org/10.1016/j.intmar.2013.12.002

Holsapple, C. W., Hsiao, S.-H., & Pakath, R. (2018). Business social media analytics: Characterization and conceptual framework. *Decision Support Systems*, *110*, 32-45. https://doi.org/10.1016/j.dss.2018.03.004

Hooley, G. J., Piercy, N., & Nicoulaud, B. (2008). *Marketing strategy and competitive positioning*. Pearson Education.

Hornik, J., Shaanan Satchi, R., Cesareo, L., & Pastore, A. (2015). Information dissemination via electronic word-of-mouth: Good news travels fast, bad news travels faster. *Computers in Human Behavior*, *45*, 273-280. https://doi.org/10.1016/j.chb.2014.11.008

Inzalkar, S., & Sharma, J. (2015). A survey on text mining-techniques and application. *International Journal of Research In Science & Engineering*, *24*, 1-14.

Ismagilova, E., Dwivedi, Y. K., Slade, E., & Williams, M. D. (2017). *Electronic Word of Mouth (eWOM) in the Marketing Context: A State of the Art Analysis and Future Directions*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-52459-7

Istanbulluoglu, D. (2017). Complaint handling on social media: The impact of multiple response times on consumer satisfaction. *Computers in Human Behavior*, *74*, 72-82. https://doi.org/10.1016/j.chb.2017.04.016

Jagtap, V., & Pawar, K. (2013). Analysis of different approaches to sentence-level sentiment classification. *International Journal of Scientific Engineering and Technology*, *2*(3), 164-170.

Jalilvand, M. R., & Heidari, A. (2017). Comparing face-to-face and electronic word-of-mouth in destination image formation: The case of Iran. *Information technology & people (West Linn, Or.)*, *30*(4), 710-735. https://doi.org/10.1108/ITP-09-2016-0204

Jan, B., Farman, H., Khan, M., Imran, M., Islam, I. U., Ahmad, A., Ali, S., & Jeon, G. (2019). Deep learning in big data Analytics: A comparative study. *Computers & electrical engineering*, *75*, 275-287. https://doi.org/10.1016/j.compeleceng.2017.12.009

Jin, J., Liu, Y., Ji, P., & Kwong, C. K. (2019). Review on recent advances in information mining from big consumer opinion data for product design. *Journal of Computing and Information Science in Engineering*, *19*(1).

Jo, T. (2018). *Text mining: Concepts, implementation, and big data challenge* (Vol. 45). Springer.

Judith, A. C., & Dina, M. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of marketing research*, *43*(3), 345-354. https://doi.org/10.1509/jmkr.43.3.345

Junni, P., Sarala, R. M., Tarba, S. Y., & Weber, Y. (2015). The Role of Strategic Agility in Acquisitions. *British journal of management*, *26*(4), 596-616. https://doi.org/10.1111/1467-8551.12115

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*.

Kao, A., & Poteet, S. R. (2007). *Natural language processing and text mining*. Springer Science & Business Media.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, *53*(1), 59-68. https://doi.org/10.1016/j.bushor.2009.09.003

Keaveney, S. M. (1995). Customer Switching Behavior in Service Industries: An Exploratory Study. *Journal of Marketing*, *59*(2), 71-82. https://doi.org/10.2307/1252074

Khan, Z., & Vorley, T. (2017). Big data text analytics: an enabler of knowledge management. *Journal of knowledge management*.

Khare, A., Labrecque, L. I., & Asare, A. K. (2011). The Assimilative and Contrastive Effects of Word-of-Mouth Volume: An Experimental Examination of Online Consumer Ratings. *Journal of retailing*, *87*(1), 111-126. https://doi.org/10.1016/j.jretai.2011.01.005

Kiecker, P., & Cowles, D. (2002). Interpersonal communication and personal influence on the Internet: A framework for examining online word-of-mouth. *Journal of Euromarketing*, *11*(2), 71-88.

Kim, K., Yoon, S., & Choi, Y. K. (2019). The effects of eWOM volume and valence on product sales - an empirical examination of the movie industry. *International journal of advertising*, *38*(3), 471-488. https://doi.org/10.1080/02650487.2018.1535225

Kim, S., Koh, Y., Cha, J., & Lee, S. (2015). Effects of social media on firm value for U.S. restaurant companies. *International Journal of Hospitality Management*, *49*, 40-46. https://doi.org/10.1016/j.ijhm.2015.05.006

Kim, W. G., Lim, H., & Brymer, R. A. (2015). The effectiveness of managing social media on hotel performance. *International Journal of Hospitality Management*, *44*, 165-171. https://doi.org/10.1016/j.ijhm.2014.10.014

Koch, C. H., & Gyrd-Jones, R. I. (2019). Corporate brand positioning in complex industrial firms: Introducing a dynamic, process approach to positioning. *Industrial marketing management*, *81*, 40-53. https://doi.org/10.1016/j.indmarman.2019.03.011

Kogut, B., & Zander, U. (1992). Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization science (Providence, R.I.)*, *3*(3), 383-397. https://doi.org/10.1287/orsc.3.3.383

Kopp, T., Riekert, M., & Utz, S. (2018). When cognitive fit outweighs cognitive load: Redundant data labels in charts increase accuracy and speed of information extraction. *Computers in Human Behavior*, *86*, 367-376. https://doi.org/10.1016/j.chb.2018.04.037

Kotsilieris, T., Pavlaki, A., Christopoulou, S., & Anagnostopoulos, I. (2017). The impact of social networks on health care. *Social network analysis and mining*, *7*(1), 1-6. https://doi.org/10.1007/s13278-017-0438-1

Kowsari, Jafari, M., Heidarysafa, Mendu, Barnes, & Brown. (2019). Text Classification Algorithms: A Survey. *Information (Basel)*, *10*(4), 150. https://doi.org/10.3390/info10040150

Krishnamurthy, B., & Wang, J. (2000). On network-aware clustering of Web clients. *Computer communication review*, *30*(4), 97-110. https://doi.org/10.1145/347057.347412

Ku, C.-H., Chang, Y.-C., Wang, Y., Chen, C.-H., & Hsiao, S.-H. (2019). Artificial intelligence and visual analytics: A deep-learning approach to analyze hotel reviews & responses.

Kunz, W., Aksoy, L., Bart, Y., Heinonen, K., Kabadayi, S., Ordenes, F. V., Sigala, M., Diaz, D., & Theodoulidis, B. (2017). Customer engagement in a Big Data world. *The Journal of services marketing*, *31*(2), 161-171. https://doi.org/10.1108/JSM-10-2016-0352

Ladhari, R., & Michaud, M. (2015). eWOM effects on hotel booking intentions, attitudes, trust, and website perceptions. *International Journal of Hospitality Management*, *46*, 36-45. https://doi.org/10.1016/j.ijhm.2015.01.010

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, *6*(70), 1.
Lappas, T., Sabnis, G., & Valkanas, G. (2016). The Impact of Fake Reviews on Online Visibility: A Vulnerability Assessment of the Hotel Industry. *Information systems research*, *27*(4), 940-961. https://doi.org/10.1287/isre.2016.0674
Laroche, M., Habibi, M. R., & Richard, M.-O. (2013). To be or not to be in social media: How brand loyalty is affected by social media? *International journal of information management*, *33*(1), 76-82. https://doi.org/10.1016/j.ijinfomgt.2012.07.003
Lee, C. H., & Cranage, D. A. (2014). Toward Understanding Consumer Processing of Negative Online Word-of-Mouth Communication: The Roles of Opinion Consensus and Organizational Response Strategies. *Journal of hospitality & tourism research (Washington, D.C.)*, *38*(3), 330-360. https://doi.org/10.1177/1096348012451455
Lee, Y. L., & Song, S. (2010). An empirical investigation of electronic word-of-mouth: Informational motive and corporate response strategy. *Computers in Human Behavior*, *26*(5), 1073-1080. https://doi.org/10.1016/j.chb.2010.03.009
Leskovec, J., Adamic, L. A., & Huberman, B. A. (2005). The Dynamics of Viral Marketing. https://doi.org/10.1145/1232722.1232727
Leung, D., Law, R., Van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of travel & tourism marketing*, *30*(1-2), 3-22.
Levy, S., & Gvili, Y. (2015). How credible is e-word of mouth across digital-marketing channels? The roles of social capital, information richness, and interactivity. *Journal of advertising research*, *55*(1), 95-109. https://doi.org/10.2501/JAR-55-1-095-109
Levy, S. E., Duan, W., & Boo, S. (2013). An Analysis of One-Star Online Reviews and Responses in the Washington, D.C., Lodging Market. *Cornell hospitality quarterly*, *54*(1), 49-63. https://doi.org/10.1177/1938965512464513
Li, C., Cui, G., & He, Y. (2020). The role of explanations and metadiscourse in management responses to anger-reviews versus anxiety-reviews: The mediation of sense-making. *International Journal of Hospitality Management*, *89*, 102560. https://doi.org/10.1016/j.ijhm.2020.102560
Li, C., Cui, G., & Peng, L. (2017). The signaling effect of management response in engaging customers: A study of the hotel industry. *Tourism management (1982)*, *62*, 42-53. https://doi.org/10.1016/j.tourman.2017.03.009
Li, C., Cui, G., & Peng, L. (2018). Tailoring management response to negative reviews: The effectiveness of accommodative versus defensive responses. *Computers in Human Behavior*, *84*, 272-284. https://doi.org/10.1016/j.chb.2018.03.009
Li, H., Bhowmick, S. S., & Sun, A. (2011). AffRank: Affinity-driven ranking of products in online social rating networks. *Journal of the American Society for Information Science and Technology*, *62*(14), 1345-1359. https://doi.org/10.1002/asi.21555
Li, J., Sun, A., Han, J., & Li, C. (2020). A survey on deep learning for named entity recognition. *IEEE Transactions on knowledge and data engineering*.
Li, K., Zhang, L., & Huang, H. (2018). Social Influence Analysis: Models, Methods, and Evaluation. *Engineering (Beijing, China)*, *4*(1), 40-46. https://doi.org/10.1016/j.eng.2018.02.004
Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, *48*(2), 354-368. https://doi.org/10.1016/j.dss.2009.09.003
Lim, B. C., & Chung, C. M. Y. (2011). The impact of word-of-mouth communication on attribute evaluation. *Journal of Business Research*, *64*(1), 18-23. https://doi.org/10.1016/j.jbusres.2009.09.014

Lim, S., Tucker, C. S., & Kumara, S. (2017). An unsupervised machine learning model for discovering latent infectious diseases using social media data. *Journal of biomedical informatics*, *66*, 82-94. https://doi.org/10.1016/j.jbi.2016.12.007

Lis, B., & Neßler, C. (2014). Electronic Word of Mouth. *Business & Information Systems Engineering*, *6*(1), 63-65. https://doi.org/10.1007/s12599-013-0306-0

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism management*, *29*(3), 458-468.

Liu, X., Schuckert, M., & Law, R. (2018). Utilitarianism and knowledge growth during status seeking: Evidence from text mining of online reviews. *Tourism management (1982)*, *66*, 38-46. https://doi.org/10.1016/j.tourman.2017.11.005

Lopez, F., Romero, V., & Sharma, G. (2014). *Mastering python regular expressions : leverage regular expressions in python even for the most complex features* (1st edition ed.). Packt Publishing Ltd.

Luo, X., & Homburg, C. (2007). Neglected outcomes of customer satisfaction. *Journal of Marketing*, *71*(2), 133-149. https://doi.org/10.1509/jmkg.71.2.133

Lyu, K., & Kim, H. (2016). Sentiment Analysis Using Word Polarity of Social Media. *Wireless personal communications*, *89*(3), 941-958. https://doi.org/10.1007/s11277-016-3346-1

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing – Connecting computing power to human insights. *International journal of research in marketing*, *37*(3), 481-504. https://doi.org/10.1016/j.ijresmar.2020.04.005

Ma, L., Sun, B., & Kekre, S. (2015). The Squeaky Wheel Gets the Grease—An Empirical Analysis of Customer Voice and Firm Intervention on Twitter. *Marketing science (Providence, R.I.)*, *34*(5), 627-645. https://doi.org/10.1287/mksc.2015.0912

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Marshall, A., Mueck, S., & Shockley, R. (2015). How leading organizations use big data and analytics to innovate. *Strategy & leadership*, *43*(5), 32-39. https://doi.org/10.1108/SL-06-2015-0054

Marti, I., & Fernandez, P. (2013). The Institutional Work of Oppression and Resistance: Learning from the Holocaust. *Organization studies*, *34*(8), 1195-1223. https://doi.org/10.1177/0170840613492078

Mauri, A. G., & Minazzi, R. (2013). Web reviews influence on expectations and purchasing intentions of hotel potential customers. *International Journal of Hospitality Management*, *34*(1), 99-107. https://doi.org/10.1016/j.ijhm.2013.02.012

Medvedev, A. N., Lambiotte, R., & Delvenne, J.-C. (2020). The anatomy of Reddit: An overview of academic research. In. Cornell University Library, arXiv.org. https://doi.org/10.1007/978-3-030-14683-2_9

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & management*, *57*(2), 103169. https://doi.org/10.1016/j.im.2019.05.004

Mikalef, P., & Pateli, A. (2017). Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA. *Journal of Business Research*, *70*, 1-16. https://doi.org/10.1016/j.jbusres.2016.09.004

Min, H., Lim, Y., & Magnini, V. P. (2015). Factors Affecting Customer Satisfaction in Responses to Negative Online Hotel Reviews: The Impact of Empathy, Paraphrasing, and Speed.

Cornell hospitality quarterly, *56*(2), 223-231. https://doi.org/10.1177/1938965514560014
Minelli, M., Chambers, M., & Dhiraj, A. (2013). *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses* (Vol. 578). John Wiley & Sons.
Minnema, A., Bijmolt, T. H. A., Gensler, S., & Wiesel, T. (2016). To Keep or Not to Keep: Effects of Online Customer Reviews on Product Returns. *Journal of retailing*, *92*(3), 253-267. https://doi.org/10.1016/j.jretai.2016.03.001
Mitra, S. K., Chattopadhyay, M., & Jana, R. K. (2019). Spillover analysis of tourist movements within Europe. *Annals of tourism research*, *79*, 102754. https://doi.org/10.1016/j.annals.2019.102754
Moe, W. W., & Schweidel, D. A. (2012). Online Product Opinions: Incidence, Evaluation, and Evolution. *Marketing science (Providence, R.I.)*, *31*(3), 372-386. https://doi.org/10.1287/mksc.1110.0662
Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.
Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, *40*(2), 621-633.
Morgan, N. A. (2011). Marketing and business performance. *Journal of the Academy of Marketing Science*, *40*(1), 102-119. https://doi.org/10.1007/s11747-011-0279-9
Murthi, B. P. S., & Sarkar, S. (2003). The Role of the Management Sciences in Research on Personalization. *Management Science*, *49*(10), 1344-1362. https://doi.org/10.1287/mnsc.49.10.1344.17313 (Management Science)
Murugesan, S. (2007). Understanding Web 2.0. *IT professional*, *9*(4), 34-41. https://doi.org/10.1109/MITP.2007.78
Musen, M. A., Bean, C. A., Cheung, K.-H., Dumontier, M., Durante, K. A., Gevaert, O., Gonzalez-Beltran, A., Khatri, P., Kleinstein, S. H., O'Connor, M. J., Pouliot, Y., Rocca-Serra, P., Sansone, S.-A., & Wiser, J. A. (2015). The center for expanded data annotation and retrieval. *Journal of the American Medical Informatics Association : JAMIA*, *22*(6), 1148-1152. https://doi.org/10.1093/jamia/ocv048
Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of big data*, *2*(1), 1-21. https://doi.org/10.1186/s40537-014-0007-7
Nenkova, A., & McKeown, K. (2011). *Automatic summarization*. Now Publishers Inc.
Nguyen, D. T., Hwang, D., & Jung, J. J. (2015). Time-frequency social data analytics for understanding social big data. In *Intelligent distributed computing VIII* (pp. 223-228). Springer.
Nieto, J., Hernández-Maestro, R. M., & Muñoz-Gallego, P. A. (2014). Marketing decisions, customer reviews, and business performance: The use of the Toprural website by Spanish rural lodging establishments. *Tourism management (1982)*, *45*, 115-123. https://doi.org/10.1016/j.tourman.2014.03.009
Nikolay, A., Anindya, G., & Panagiotis, G. I. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, *57*(8), 1485-1509. https://doi.org/10.1287/mnsc.1110.1370 (Management Science)
Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization science (Providence, R.I.)*, *5*(1), 14-37. https://doi.org/10.1287/orsc.5.1.14
O’reilly, T. (2005). Web 2.0: compact definition. In.

Obeidat, Z. M. I., Xiao, S. H., Iyer, G. R., & Nicholson, M. (2017). Consumer Revenge Using the Internet and Social Media: An Examination of the Role of Service Failure Types and Cognitive Appraisal Processes: CONSUMER REVENGE USING ONLINE AND SOCIAL MEDIA. *Psychology & marketing*, *34*(4), 496-515. https://doi.org/10.1002/mar.21002

Okazaki, S., & Taylor, C. R. (2013). Social media and international advertising: theoretical challenges and future directions. *International marketing review*, *30*(1), 56-71. https://doi.org/10.1108/02651331311298573

Palese, B., & Usai, A. (2018). The relative importance of service quality dimensions in E-commerce experiences. *International journal of information management*, *40*, 132-140. https://doi.org/10.1016/j.ijinfomgt.2018.02.001

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345-1359.

Park, S.-Y., & Allen, J. P. (2013). Responding to Online Reviews: Problem Solving and Engagement in Hotels. *Cornell hospitality quarterly*, *54*(1), 64-73. https://doi.org/10.1177/1938965512463118

Pathak, A. R., Pandey, M., & Rautaray, S. (2019). Adaptive Model for Dynamic and Temporal Topic Modeling from Big Data using Deep Learning Architecture. *International journal of intelligent systems and applications*, *11*(6), 13-27. https://doi.org/10.5815/ijisa.2019.06.02

Patil, S., & Kulkarni, S. (2018). Mining Social Media Data for Understanding Students' Learning Experiences using Memetic algorithm. *Materials Today: Proceedings*, *5*(1), 693-699. https://doi.org/10.1016/j.matpr.2017.11.135

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.

Pawar, S., Palshikar, G. K., & Bhattacharyya, P. (2017). Relation extraction: A survey. *arXiv preprint arXiv:1712.05191*.

Peng, S., Wang, G., & Xie, D. (2017). Social Influence Analysis in Social Networking Big Data: Opportunities and Challenges. *IEEE network*, *31*(1), 11-17. https://doi.org/10.1109/MNET.2016.1500104NM

Peppers, D., & Rogers, M. (1993). *The one to one future: Building relationships one customer at a time*. Currency Doubleday New York.

Piehler, R., Schade, M., Hanisch, I., & Burmann, C. (2019). Reacting to negative online customer reviews: Effects of accommodative management responses on potential customers. *Journal of service theory and practice*, *29*(4), 401-414. https://doi.org/10.1108/JSTP-10-2018-0227

Plotkina, D., & Munzel, A. (2016). Delight the experts, but never dissatisfy your customers! A multi-category study on the effects of online review source on intention to buy a new product. *Journal of retailing and consumer services*, *29*, 1-11. https://doi.org/10.1016/j.jretconser.2015.11.002

Proserpio, D., & Zervas, G. (2017). Online Reputation Management: Estimating the Impact of Management Responses on Consumer Reviews. *Marketing science (Providence, R.I.)*, *36*(5), 645-665. https://doi.org/10.1287/mksc.2017.1043

Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. " O'Reilly Media, Inc.".

Rahmani, A., Chen, A., Sarhan, A., Jida, J., Rifaie, M., & Alhajj, R. (2014). Social media analysis and summarization for opinion mining: a business case study. *Social network analysis and mining*, *4*(1), 1-11. https://doi.org/10.1007/s13278-014-0171-y

Raj, A., Adam, R., Prabakar, K., & Rakesh, K. S. (2012). An Emotion-Based Model of Salesperson Ethical Behaviors. *Journal of Business Ethics*, *109*(2), 243-257. https://doi.org/10.1007/s10551-011-1123-3
Ramchand, G., & Reiss, C. (2007). *The Oxford handbook of linguistic interfaces*. Oxford University Press.
Reimer, T., & Benkenstein, M. (2016). Altruistic eWOM marketing: More than an alternative to monetary incentives. *Journal of retailing and consumer services*, *31*, 323-333. https://doi.org/10.1016/j.jretconser.2016.04.003
Reuter, T., & Cimiano, P. (2012). Event-based classification of social media streams. Proceedings of the 2nd ACM International Conference on Multimedia Retrieval,
Rob, K. (2014). Big Data. In (pp. 67). London: London: SAGE Publications Ltd.
Robert, M. G. (1996). Toward a Knowledge-Based Theory of the Firm. *Strategic management journal*, *17*(S2), 109-122. https://doi.org/10.1002/smj.4250171110
Rodriguez, M., Peterson, R. M., & Krishnan, V. (2012). Social media's influence on business-to-business sales performance. *Journal of Personal Selling & Sales Management*, *32*(3), 365-378.
Rosenmayer, A., McQuilken, L., Robertson, N., & Ogden, S. (2018). Omni-channel service failures and recoveries: refined typologies using Facebook complaints. *The Journal of services marketing*, *32*(3), 269-285. https://doi.org/10.1108/JSM-04-2017-0117
Rydning, D. R. J. G. J. (2018). The digitization of the world from edge to core. *Framingham: International Data Corporation*.
Saha, B., & Srivastava, D. (2014). Data quality: The other face of big data. 2014 IEEE 30th International Conference on Data Engineering,
Salganik, M. J., & Watts, D. J. (2008). Leading the Herd Astray: An Experimental Study of Self-fulfilling Prophecies in an Artificial Cultural Market. *Social psychology quarterly*, *71*(4), 338-355. https://doi.org/10.1177/019027250807100404
Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Pearson education.
Saunders, M. N. K., Lewis, P., & Thornhill, A. (2019). *Research methods for business students* (Eighth edition. ed.). Pearson Education Limited.
Saxena, A., & Khanna, U. (2013). Advertising on Social Network Sites: A Structural Equation Modelling Approach. *Vision: The Journal of Business Perspective*, *17*(1), 17-25. https://doi.org/10.1177/0972262912469560
Schouten, K., & Frasincar, F. (2015). Survey on aspect-level sentiment analysis. *IEEE Transactions on knowledge and data engineering*, *28*(3), 813-830.
Scott, J., & Carrington, P. J. (2011). *The SAGE Handbook of Social Network Analysis*. London: SAGE Publications.
Sebei, H., Hadj Taieb, M. A., & Ben Aouicha, M. (2018). Review of social media analytics process and Big Data pipeline. *Social network analysis and mining*, *8*(1), 1-28. https://doi.org/10.1007/s13278-018-0507-0
Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the Web. *Journal of Interactive Marketing*, *21*(4), 76-94. https://doi.org/10.1002/dir.20090
Settles, B. (2009). *Active learning literature survey*.
Shamim, S., Zeng, J., Shariq, S. M., & Khan, Z. (2019). Role of big data management in enhancing big data decision-making capability and quality among Chinese firms: A dynamic capabilities view. *Information & management*, *56*(6), 103135. https://doi.org/10.1016/j.im.2018.12.003

Shan, S., Luo, Y., Zhou, Y., & Wei, Y. (2019). Big data analysis adaptation and enterprises' competitive advantages: the perspective of dynamic capability and resource-based theories. *Technology analysis & strategic management*, *31*(4), 406-420. https://doi.org/10.1080/09537325.2018.1516866

Sharma, S., & Verma, H. (2018). Social media marketing: Evolution and change. In *Social media marketing* (pp. 19-36). Springer.

Sheng, J. (2018). *Managing big data from the crowd: strategic firm engagement with online social interactions* University of Bristol].

Sheng, J. (2019). Being Active in Online Communications: Firm Responsiveness and Customer Engagement Behaviour. *Journal of Interactive Marketing*, *46*, 40-51. https://doi.org/10.1016/j.intmar.2018.11.004

Sheng, J., Amankwah-Amoah, J., & Wang, X. (2019). Technology in the 21st century: New challenges and opportunities. *Technological forecasting & social change*, *143*, 321-335. https://doi.org/10.1016/j.techfore.2018.06.009

Sheng, J., Amankwah-Amoah, J., Wang, X., & Khan, Z. (2019). Managerial Responses to Online Reviews: A Text Analytics Approach. *British journal of management*, *30*(2), 315-327. https://doi.org/10.1111/1467-8551.12329

Shin, D. (2020). Enhancing social media analysis with visual data analytics: a deep learning approach. *Mis Quarterly*, *44*(4), 1459-1492. https://doi.org/10.25300/MISQ/2020/14870

Sorescu, A. (2017). Data-Driven Business Model Innovation. *The Journal of product innovation management*, *34*(5), 691-696. https://doi.org/10.1111/jpim.12398

Sparks, B. A., & Bradley, G. L. (2017). A "Triple A" Typology of Responding to Negative Consumer-Generated Online Reviews. *Journal of hospitality & tourism research (Washington, D.C.)*, *41*(6), 719-745. https://doi.org/10.1177/1096348014538052

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism management (1982)*, *32*(6), 1310-1323. https://doi.org/10.1016/j.tourman.2010.12.011

Sparks, B. A., So, K. K. F., & Bradley, G. L. (2016). Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern. *Tourism management (1982)*, *53*, 74-85. https://doi.org/10.1016/j.tourman.2015.09.011

Stieglitz, S., & Dang-Xuan, L. (2012). Social media and political communication: a social media analytics framework. *Social network analysis and mining*, *3*(4), 1277-1291. https://doi.org/10.1007/s13278-012-0079-3

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International journal of information management*, *39*, 156-168. https://doi.org/10.1016/j.ijinfomgt.2017.12.002

Stone, J. E., Perilla, J. R., Cassidy, C. K., & Schulten, K. (2017). GPU-accelerated molecular dynamics clustering analysis with OpenACC. In *Parallel Programming with OpenACC* (pp. 215-240). Elsevier.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Talabis, M., McPherson, R., Miyamoto, I., & Martin, J. (2014). *Information Security Analytics: Finding Security Insights, Patterns, and Anomalies in Big Data*. Syngress.

Tan, L. K.-W., Na, J.-C., Theng, Y.-L., & Chang, K. (2011). Sentence-level sentiment polarity classification using a linguistic approach. International Conference on Asian Digital Libraries,

Tang, D. (2015). Sentiment-specific representation learning for document-level sentiment analysis. Proceedings of the eighth ACM international conference on web search and data mining, Teece, D., Peteraf, M., & Leih, S. (2016). Dynamic Capabilities and Organizational Agility: Risk, Uncertainty, and Strategy in the Innovation Economy. *California management review*, *58*(4), 13-35. https://doi.org/10.1525/cmr.2016.58.4.13

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic management journal*, *28*(13), 1319-1350. https://doi.org/10.1002/smj.640

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal*, *18*(7), 509-533.

Tess, P. A. (2013). The role of social media in higher education classes (real and virtual) – A literature review. *Computers in Human Behavior*, *29*(5), A60-A68. https://doi.org/10.1016/j.chb.2012.12.032

Thomas, H. D. (2013). Analytics 3.0. *Harvard business review*, *91*(12), 64.

Thumvichit, A., & Gampper, C. (2019). Composing responses to negative hotel reviews: A genre analysis. *Cogent arts & humanities*, *6*(1). https://doi.org/10.1080/23311983.2019.1629154

Tian, X. (2017). Big data and knowledge management: a case of déjà vu or back to the future? *Journal of knowledge management*, *21*(1), 113-131. https://doi.org/10.1108/JKM-07-2015-0277

TripAdvisor. (2018). *Media Center*. Retrieved 07 April 2019 from https://tripadvisor.mediaroom.com

Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of Word-of-Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site. *Journal of Marketing*, *73*(5), 90-102. https://doi.org/10.1509/jmkg.73.5.90

Tsao, W.-C., & Hsieh, M.-T. (2015). eWOM persuasiveness: do eWOM platforms and product type matter? *Electronic Commerce Research*, *15*(4), 509-541. https://doi.org/10.1007/s10660-015-9198-z

Tuten, T. L. (2018). *Social media marketing / Tracy L. Tuten, Michael R. Solomon* (3rd edition. ed.). London : SAGE Publications Ltd.

Umashankar, N., Ward, M. K., & Dahl, D. W. (2017). The Benefit of Becoming Friends: Complaining after Service Failures Leads Customers with Strong Ties to Increase Loyalty. *Journal of Marketing*, *81*(6), 79-98. https://doi.org/10.1509/jm.16.0125

Van Der Maaten, L., Postma, E., & Van den Herik, J. (2009). Dimensionality reduction: a comparative. *J Mach Learn Res*, *10*(66-71), 13.

Vatrapu, R., Mukkamala, R. R., Hussain, A., & Flesch, B. (2016). Social Set Analysis: A Set Theoretical Approach to Big Data Analytics. *IEEE access*, *4*, 2542-2571. https://doi.org/10.1109/ACCESS.2016.2559584

Vikas, A., William, H. G., & Charles, C. M. (2002). Thriving on the Knowledge of Outsiders: Tapping Organizational Social Capital. *The Academy of Management executive (1993)*, *16*(1), 87-101. https://doi.org/10.5465/AME.2002.6640198

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356-365. https://doi.org/10.1016/j.jbusres.2016.08.009

Wan, F., & Ren, F. (2017). The effect of firm marketing content on product sales: Evidence from a mobile social media platform. *Journal of electronic commerce research*, *18*(4), 288-302.

Wang, F., Liu, X., & Fang, E. (2015). User Reviews Variance, Critic Reviews Variance, and Product Sales: An Exploration of Customer Breadth and Depth Effects. *Journal of retailing*, *91*(3), 372-389. https://doi.org/10.1016/j.jretai.2015.04.007

Wang, H., Lu, Y., & Zhai, C. (2010). Latent aspect rating analysis on review text data: a rating regression approach. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining,

Wang, Y.-S., Wu, S.-C., Lin, H.-H., & Wang, Y.-Y. (2011). The relationship of service failure severity, service recovery justice and perceived switching costs with customer loyalty in the context of e-tailing. *International journal of information management*, *31*(4), 350-359. https://doi.org/10.1016/j.ijinfomgt.2010.09.001

Wang, Y., & Chaudhry, A. (2018). When and how Managers' Responses to Online Reviews Affect Subsequent Reviews. *Journal of marketing research*, *55*(2), 163-177. https://doi.org/10.1509/jmr.15.0511

Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016). Attention-based LSTM for aspect-level sentiment classification. Proceedings of the 2016 conference on empirical methods in natural language processing,

Wang, Z., & Kim, H. G. (2017). Can Social Media Marketing Improve Customer Relationship Capabilities and Firm Performance? Dynamic Capability Perspective. *Journal of Interactive Marketing*, *39*, 15-26. https://doi.org/10.1016/j.intmar.2017.02.004

Weber, Y., & Tarba, S. Y. (2014). Strategic Agility: A State of the Art Introduction to the Special Section on Strategic Agility. *California management review*, *56*(3), 5-12. https://doi.org/10.1525/cmr.2014.56.3.5

Wei, W., Miao, L., & Huang, Z. (2013). Customer engagement behaviors and hotel responses. *International Journal of Hospitality Management*, *33*, 316-330. https://doi.org/10.1016/j.ijhm.2012.10.002

Weisfeld-Spolter, S., Sussan, F., & Gould, S. (2014). An integrative approach to eWOM and marketing communications. *Corporate communications*, *19*(3), 260-274. https://doi.org/10.1108/CCIJ-03-2013-0015

Wessel, M. (2016). How big data is changing disruptive innovation. *Harvard business review*, *27*.

White, M. (2012). Digital workplaces: Vision and reality. *Business information review*, *29*(4), 205-214. https://doi.org/10.1177/0266382112470412

White, T. (2015). *Hadoop: The Definitive Guide*. O'Reilly.

Wiering, M., & Van Otterlo, M. (2012). *Reinforcement learning* (Vol. 12). Springer.

Wu, F., & Huberman, B. A. (2008). How public opinion forms. International Workshop on Internet and Network Economics,

Xie, K., Kwok, L., & Wang, W. (2017). Monetizing Managerial Responses on TripAdvisor: Performance Implications Across Hotel Classes. *Cornell hospitality quarterly*, *58*(3), 240-252. https://doi.org/10.1177/1938965516686109

Xie, K. L., So, K. K. F., & Wang, W. (2017). Joint effects of management responses and online reviews on hotel financial performance: A data-analytics approach. *International Journal of Hospitality Management*, *62*, 101-110. https://doi.org/10.1016/j.ijhm.2016.12.004

Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, *43*, 1-12. https://doi.org/10.1016/j.ijhm.2014.07.007

Xie, K. L., Zhang, Z., Zhang, Z., Singh, A., & Lee, S. K. (2016). Effects of managerial response on consumer eWOM and hotel performance: Evidence from TripAdvisor. *International journal of contemporary hospitality management*, *28*(9), 2013-2034. https://doi.org/10.1108/IJCHM-06-2015-0290

Xu, F., Pan, Z., & Xia, R. (2020). E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. *Information processing & management*, *57*(5), 102221. https://doi.org/10.1016/j.ipm.2020.102221

Xu, H., Guo, H., Zhang, J., & Dang, A. (2018). Facilitating dynamic marketing capabilities development for domestic and foreign firms in an emerging economy. *Journal of Business Research*, *86*, 141-152. https://doi.org/10.1016/j.jbusres.2018.01.038

Xue-Wen, C., & Xiaotong, L. (2014). Big Data Deep Learning: Challenges and Perspectives. *IEEE access*, *2*, 514-525. https://doi.org/10.1109/access.2014.2325029

Yang, F., Zhong, B., Kumar, A., Chow, S.-M., & Ouyang, A. (2018). Exchanging Social Support Online: A Longitudinal Social Network Analysis of Irritable Bowel Syndrome Patients' Interactions on a Health Forum. *Journalism & mass communication quarterly*, *95*(4), 1033-1057. https://doi.org/10.1177/1077699017729815

Yang, S., Lin, S., Carlson, J. R., & Ross, W. T. (2016). Brand engagement on social media: will firms' social media efforts influence search engine advertising effectiveness? *Journal of marketing management*, *32*(5-6), 526-557. https://doi.org/10.1080/0267257X.2016.1143863

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies,

Yasmeen, R. (2019). *Top 100 City Destinations*
2019 Edition. E. International. https://gate.ahram.org.eg/Media/News/2019/12/9/2019-637114875902629480-262.pdf

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*(1), 180-182. https://doi.org/10.1016/j.ijhm.2008.06.011

Yoo, C. W., Sanders, G. L., & Moon, J. (2013). Exploring the effect of e-WOM participation on e-Loyalty in e-commerce. *Decision Support Systems*, *55*(3), 669-678. https://doi.org/10.1016/j.dss.2013.02.001

Zablocki, A., Schlegelmilch, B., & Houston, M. J. (2018). How valence, volume and variance of online reviews influence brand attitudes. *AMS review*, *9*(1-2), 61-77. https://doi.org/10.1007/s13162-018-0123-1

Zaobo, H., Zhipeng, C., & Xiaoming, W. (2015). Modeling Propagation Dynamics and Developing Optimized Countermeasures for Rumor Spreading in Online Social Networks. In (pp. 205-214): IEEE.

Zeng, D., Hsinchun, C., Lusch, R., & Shu-Hsing, L. (2010). Social Media Analytics and Intelligence. *IEEE intelligent systems*, *25*(6), 13-16. https://doi.org/10.1109/MIS.2010.151

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information fusion*, *42*, 146-157. https://doi.org/10.1016/j.inffus.2017.10.006

Zhang, X., Qiao, S., Yang, Y., & Zhang, Z. (2020). Exploring the impact of personalized management responses on tourists' satisfaction: A topic matching perspective. *Tourism management (1982)*, *76*, 103953. https://doi.org/10.1016/j.tourman.2019.103953

Zhang, Y., & Vásquez, C. (2014). Hotels' responses to online reviews: Managing consumer dissatisfaction. *Discourse, context & media*, *6*, 54-64. https://doi.org/10.1016/j.dcm.2014.08.004

Zhang, Z., Ye, Q., & Law, R. (2011). Determinants of hotel room price: An exploration of travelers' hierarchy of accommodation needs. *International journal of contemporary hospitality management*, *23*(7), 972-981. https://doi.org/10.1108/09596111111167551

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, *115*, 213-237.

Zhou, C., Sun, C., Liu, Z., & Lau, F. (2015). A C-LSTM neural network for text classification. *arXiv preprint arXiv:1511.08630*.
Zhu, F., & Zhang, X. (2006). The influence of online consumer reviews on the demand for experience goods: The case of video games. *ICIS 2006 Proceedings*, 25.
Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on Sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, *74*(2), 133-148. https://doi.org/10.1509/jmkg.74.2.133
Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, *3*(1), 1-130.
Zikopoulos, P. C., Deroos, D., & Parasuraman, K. (2013). *Harness the power of big data: The IBM big data platform*. McGraw-Hill.
Zirn, C., Niepert, M., Stuckenschmidt, H., & Strube, M. (2011). Fine-grained sentiment analysis with structural features. Proceedings of 5th International Joint Conference on Natural Language Processing,