Utilizing social media to the benefit of companies

L Botha
22206914

Dissertation submitted in partial fulfilment of the requirements for the degree *Magister Scientiae* in Computer Science at the Potchefstroom Campus of the North-West University

Supervisor: Prof HM Huisman
Co-supervisor: Dr E Taylor

May 2017

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

It all starts here ™

ACKNOWLEDGEMENTS

"God never said that the journey would be easy, but He did say that the arrival would be worthwhile." – Max Lucado

ABSTRACT

The use of social media in companies is a complex subject, which requires more research. The knowledge of companies regarding the use of social media is limited and companies are often faced with the problem of choosing techniques, strategies and frameworks that can be utilized to the benefit of the company. Social media can be used for a variety of purposes, for example to increase brand awareness, as a marketing tool, to improve the engagement between the company and its customers, etc. The aim of this study is to identify which techniques a company can use to ensure that social media is utilized effectively.

For this study, a positivist research paradigm was adopted to ensure that the researcher has an objective view of the reality. The quality of this positivistic study was assessed according to the following criteria: objectivity, reliability, internal validity and external validity. This research paradigm was chosen because it is regarded as the dominant paradigm when researching information systems, computer science or information technology.

To identify which techniques, strategies and frameworks are currently being used to manage social media of a company, a survey was conducted. The data generation method chosen to complement the survey research method was web-based questionnaires. A total of 122 questionnaires were returned from employees in different departments and different companies. Twitter data of three different South African companies was collected and analysed. A tweet was categorised according to positive or negative, indicating the sentiment, as well as subjective or objective, indicating the opinion. Sentiment analysis and opinion mining offers a company the ability to observe a variety of social media platforms in real time and to act accordingly in respect of the social media data gathered.

Each company is unique, not only the type of company, but also the company's customers or clients. When choosing a technique to ensure that social media is utilized effectively, the techniques should be chosen in such a manner that it is fitting to what the company wants to achieve by using social media. If a company wants to measure the activity and awareness of customers/clients on the company's social media platforms, techniques that can be used include reporting tools, advertisement managing software, Application Programming Interfaces (API's) of different social media platforms, social media management systems (for example Hootsuite) and built-in dashboards offered by social media platforms. A basic framework was developed and can be used by companies as a tool or guideline when deciding to use social media.

OPSOMMING

Die gebruik van sosiale media in maatskappye is 'n komplekse onderwerp wat meer navorsing vereis. Die kennis van maatskappye met betrekking tot die gebruik van sosiale media is beperk en maatskappye word dikwels gekonfronteer met die probleem om 'n keuse te maak tussen verskillende tegnieke, strategieë en sosiale mediaraamwerke wat gebruik kan word tot voordeel van die maatskappy. Sosiale media kan gebruik word vir 'n verskeidenheid van doeleindes, byvoorbeeld om bewustheid van 'n produk of diens te verhoog, as 'n bemarkingsinstrument, om die betrokkenheid tussen die maatskappy en hul kliënte te verbeter, ens. Die doel van hierdie studie is om te bepaal watter tegnieke 'n maatskappy kan gebruik om te verseker dat sosiale media effektief benut word.

Vir hierdie studie is 'n positivistiese navorsingsparadigma gevolg om te verseker dat die navorser 'n objektiewe siening van die werklikheid behou. Die kwaliteit van hierdie positivistiese studie is beoordeel volgens die volgende kriteria: objektiwiteit, betroubaarheid, interne geldigheid en eksterne geldigheid. Dit navorsingsparadigma is gekies omdat dit beskou word as die dominante paradigma vir die ondersoek van inligtingstelsels, rekenaarwetenskap of inligtingstegnologie.

Om te bepaal watter tegnieke, strategieë en raamwerke tans gebruik word om sosiale media van 'n maatskappy te bestuur, is daar van 'n opname gebruik gemaak. 'n Web-gebaseerde vraelys is gebruik as datagenereringsmetode. 'n Totaal van 122 vraelyste is terug ontvang van werknemers in verskillende departemente in verskillende maatskappye. Twitter-data van drie verskillende Suid-Afrikaanse maatskappye is ingesamel en ontleed. 'n Toet is gekategoriseer as positief of negatief, wat die sentiment aandui, sowel as subjektief of objektief, wat die mening van die toet aandui. Sentimentanalise en die ontginning van menings bied 'n maatskappy die vermoë om 'n verskeidenheid van sosiale mediaplatforms in reële tyd te ondersoek en op te tree volgens die sosiale media data wat dan ingesamel is.

Elke maatskappy is uniek, nie net die tipe maatskappy nie, maar ook die maatskappy se kliënte. Wanneer 'n tegniek gekies word om te verseker dat sosiale media effektief gebruik word, moet die tegnieke op sodanige wyse gekies word dat dit gepas is by dit wat die maatskappy wil bereik deur die gebruik van sosiale media. Indien 'n maatskappy die aktiwiteit en bewustheid van kliënte wil bepaal op die maatskappy se sosiale media-platforms, kan die volgende tegnieke gebruik word: verslagdoeningshulpmiddels, advertensiebestuursagteware, toepassingprogammeringskoppelvlakke (API's) van verskillende sosiale mediaplatforms, sosiale mediabestuurstelsels (byvoorbeeld Hootsuite) en ingeboude sagtewarepaneelborde wat verkry kan word op sosiale mediaplatforms. 'n Basiese raamwerk is ontwikkel en kan deur

maatskappye gebruik word as 'n instrument of riglyn wanneer besluit word om sosiale media te gebruik.

Sleutelterme: sosiale media, kliënteverhoudingsbestuur, sosiale mediaplatforms, sosiale media data, sentimentanalise, meningsontginning, data-ontginning

TABLE OF CONTENTS:

LIST OF TABLES:

LIST OF FIGURES:

CHAPTER 1: PROBLEM STATEMENT

1.1 Introduction

The aim of this chapter is to provide an introduction regarding the particulars of this study. In section 1.2 the problem statement is discussed, the research question is stated and areas to which the research will contribute are discussed. In section 1.3 previous studies done on the use of social media in companies are discussed. The research aims and objectives are discussed in section 1.4. The research aims and objectives will be used as guidelines in the process of answering the research question. The method of investigation is discussed in section 1.5. In section 1.6 the outline of chapters is given. Figure 1.1 is a representation of what will be discussed in this chapter in particular and the study in general.

Utilizing social media to the benefit of companies

Chapter 1: Problem statement

1.1 Introduction

1.2 Problem statement and substantiation
The theme of this research is linked to literature. The research question and how the research will be conducted to answer the question are discussed.

1.3 Previous studies
Recent research relating to this topic is identified.

1.4 Research aims and objectives
The main objective is to determine which techniques can be used by companies to effectively utilize social media. The general, as well as the specific aspects that will form part of the research are also discussed.

1.5 Method of investigation
The proposed design, data acquisition, data processing, etc. are discussed.

1.6 Outline of this study

1.7 Conclusion

Chapter 2 - Literature Study: Social media	Chapter 3 - Literature study: Mining social media data

Chapter 4: Research design

Chapter 5: Results of questionnaire	Chapter 6: Results of Twitter data

Chapter 7: Discussion, interpretation and conclusion

Figure 1.1 - Overview of Chapter 1.

1.2 Problem statement and substantiation

According to statistics and research done by We Are Social (2016) 24% of South Africa's population are active social media users. This means the total number of active social media users equals a projected 13 million in South Africa. The age of social media users ranges from 13 years of age to 60+. The age group that spends the most time on social media is people aged from 20 to 29 (We Are Social, 2016). An active social media user from South Africa spend 2.7 hours each day on using social media platforms (We Are Social, 2016).

Van Zyl (2015) discusses a study done by World Wide Worx and Fuseware's South African Social Media Landscape 2016. This study revealed that the use of social media platforms has increased (Van Zyl, 2015). A survey was done during this study featuring 65 of the biggest brands in South Africa. The survey discovered that most of the major brands are currently using Facebook and Twitter. The number of people making use of social media is continuously increasing, especially in businesses (Lehtimäki *et al.,* 2009). Social media can be used internally or externally in a company.

When a company uses social media internally (Enterprise 2.0), social media technologies are applied to improve collaboration, to increase efficiency, and to encourage innovation within the company (Lardi & Fuchs, 2013:47). If social media is used externally (Business 2.0) in a company, external social media platforms are used for marketing, customer relationship management and brand awareness (Lardi & Fuchs, 2013:23). Many businesses have adopted social media as a way to enhance customer relationship management, and to meet communications objectives and business goals (He *et al.,* 2013:464; Jeffrey, 2013:2; Kaplan & Haenlein, 2010:64).

Businesses hire social media experts and consultants to determine which content, characteristics and activities in a social media environment will be efficient (Erdoğmuş & Çiçek, 2012:1355). The amount of time companies spend using social media cannot always be connected to how effective the social media is in increasing the marketing and branding strategy, as well as building relationships with customers (Paniagua & Sapena, 2014:723). Research done shows that consumers visit social media sites to keep up to date with a specific brand's products and promotional campaigns (Mangold & Faulds, 2009:357; Leggat, 2010).

Consumers observe social media sites as a service channel, where engagement between a business and consumers can occur on a real-time basis (Leggat, 2010). A company can possibly gain advantage over another company by analyzing publicly available social media data (He *et al.,* 2013:469). By analyzing social media data, a company can then redesign its processes to ensure that the company is not only brand-driven but also customer-driven. Viral marketing can be accelerated, and trend analysis, as well as sales predictions can be undertaken (Gundecha & Liu, 2012:14).

Companies struggle with identifying meaningful objectives for identifying and using social media platforms and then measuring the effectiveness of social media communication

campaigns (AMEC, 2014; Hanna *et al.,* 2011:265).

The following list is compiled from the sources listed and consists of challenges that companies face when using social media platforms (Hill, 2015; Lee, 2015; Torr, 2014):

- limited time to manage social media;
- identifying a social media platform that is suitable and effective;
- uncertain of what social media engagement is and how to ensure proactive social media engagement;
- uncertain about the profile of the user of a certain social media platform;
- uncertain about whether social media management should be outsourced;
- hesitant about the steps that should be performed to ensure that a response to a social media comment is suitable;
- cautious about which activities to post and follow; and
- uncertain about what time of day is the best to reach customers/consumers and what time is peak performance of social media traffic.

By using social media data and data-mining techniques, sentiment analysis and opinion mining can be done to determine a customer's/client's response regarding a product, brand or service. Sentiment analysis forms an important part of text mining and makes it possible to examine the authors' opinions and to determine the overall opinion of a large number of people (Dickinson & Hu, 2015:61). The aim of sentiment analysis and opinion mining is to automatically extract opinions that are expressed within the user-generated content (Gundecha & Liu, 2012:5). Although some people refer to text mining as data-mining, a clear distinction can be made between text mining, data-mining and web mining.

According to Al-Azmi (2013:2), text mining deals with textual data rather than records that can be found in a database. Text mining automatically discovers hidden patterns and unknown information. Text mining is related to data-mining and makes use of different data-mining techniques and methods. As a result of social media platforms becoming more popular by the day, these social media platforms carry huge quantities of data, which is ideal for social research purposes (Ahmed, 2015).

A great quantity of data can be collected from social media platforms and social media data is often classified as big data (Gundecha & Liu, 2012:14). Big data lack a clear and consistent definition. Jacobs (2009:44) defines big data as data of which the size forces data analysts to explore other analysis and interpretation methods than the tried-and-trusted methods.

From the above, it can be seen that South Africa has a large number of social media users, spending hours every day on social media. Companies can gain advantage by analyzing the social media data. This can be done, for example, by redesigning its processes to ensure that the company is customer-driven, accelerating viral marketing, doing trend analysis and sales prediction. However, companies struggle with the many challenges of social media platforms, for example measuring the effectiveness of social media communication campaigns.

The purpose of this study is to determine which techniques can be used by a company to ensure that social media is utilized effectively. This brings us to the research question for this study: Which techniques can be used by a company to ensure that social media is utilized effectively?

Other sources and previous studies have also been reviewed to ensure that the research question truly addresses the problem. In the next section, previous similar studies are briefly described.

1.3   Previous studies

Table 1.1 presents the researcher's own summary of similar studies that have been done regarding the use of social media in companies. The table displays the study title, findings and conclusions that have been drawn, as well as the name(s) of the author(s). The studies named in this table are based on research done from 2011 to 2015.

Table 1.1 - Previous studies

Study title	Brief description of the study	Author(s)
Using Twitter as a data source: An overview of current social media research tools	The number of tools available that can be used to obtain data from social media platforms, such as Facebook, LinkedIn, Google+, etc. is limited. Twitter data can be obtained and analyzed by using techniques, such as sentiment analysis, time series analysis, network analysis and machine-learning methods.	Ahmed (2015)
Sentiment Analysis of Investor Opinions on Twitter	Sentiment classification and sentiment correlation analysis were used to study the correlation between Twitter sentiment and the stock price. This can then be used by a company to predict the stock price. Each company used in the study had a different correlation at the end, indicating that each company and the consumers of the company are unique.	Dickinson & Hu (2015:69)
We're all connected: The power of the social media ecosystem	Platforms, such as Facebook, Twitter, YouTube, etc. recreated social media platforms to not only provide information, but also act as an influencer. One of the key areas of this study focused on ensuring rich,	Hanna *et al.,* (2011:272)

Study title	Brief description of the study	Author(s)
	meaningful and interactive dialogue with consumers.	
Social media? Get serious! Understanding the functional building blocks of social media	This study investigated whether firms should develop strategies for monitoring, understanding and responding to different social media activities. The authors of this study create a honeycomb framework with seven building blocks that companies can use as a tool when managing social media platforms.	Kietzmann, J.H. *et al.,* (2011:250)
Social Media: The Business Benefits May Be Enormous, But Can the Risks – Reputational, Legal, Operational – Be Mitigated?	In this article the benefits and risks a company should consider when using social media is discussed. Risks, such as reputational damage, employment risks, security risks, privacy risks, etc. are explained.	Merril *et al.,* (2011:6)
A Survey of Data-Mining Techniques for Social Media Analysis	Data-mining techniques have been proven useful and effective for analysing social media data and extracting opinions/sentiments. Data-mining techniques allow data scientists to work with noisy and large amounts of dynamic data. Support Vector Machines, Naïve Bayes and Maximum Entropy are the most popular techniques used for mining social media data.	Adedoyin-Olowe *et al.,* (2013)
Business engagement on Twitter: a path analysis	This study gathered information regarding the engagement of consumers and a specific brand. The results of the study were that business engagement on Twitter relates directly to consumers' engagement with on-line word-of-mouth communication. This study also revealed that the life cycle of a tweet ranges between 1.5 and 4 hours.	Zhang *et al.* (2011:173)
Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM	The results of this study indicate that social media technology usage and customer-centric management systems contribute to a firm-level capability of social customer relationship management.	Trainor *et al.* (2014)

From previous studies done, summarized in Table 1.1, there is an indication that the research regarding the use of social media within and by companies is still in the early stages. Kärkkäinen *et al.* (2010) indicate that very few academic studies have been done regarding the adoption of social media in organizations. Previous studies do not indicate on which techniques, possible frameworks or how companies can make use of social media to improve customer relationship management.

This study and research will contribute to the following areas:

- improve knowledge regarding the value of social media if managed correctly;
- provide companies with a basic framework and available techniques that can be used to analyse customers/clients feelings and paradigms about the company which can then be used to improve customer relationship management; and
- a contribution to research being done in the particular area of social media.

In order to find an answer to the main research question stated in §1.2, and to ensure that the study will contribute to the above-mentioned areas, secondary aims and objectives are set. These aims and objectives are discussed in the next section.

1.4   Research aims and objectives

The main objective of this study is to determine which techniques can be used by companies to effectively utilize social media.

To accomplish this objective the researcher will have to achieve the following secondary objectives (SO):

- SO1: Review literature on the use of social media in companies.
- SO2: Research different data- and text mining techniques that can be used to analyse social media data and to gather social media data.
- SO3: Gather and collect non-structured data, for instance "tweets" of customers from a certain company or within a certain business area.
- SO4: Determine patterns, such as identification of topic keywords, which can be found on a company's social media sites, as well as user-generated content.
- SO5: Use data- and text mining techniques and methods (identified in S02) to determine sentiment and extract hidden information.
- SO6: Create a connection between text sentiment and public opinion. This will be done by determining if a user's message articulates a negative or positive opinion regarding a certain company, brand, product or service.
- SO7: Determine how social media is used in companies today by developing a questionnaire.
- SO8: Create a basic social media framework to facilitate the analyses of social media data that can be gathered from companies' social media platforms. The framework is not a standard that should be followed but rather a recommended approach and tool.
- SO9: Draw a conclusion on the effectiveness of utilising social media in a company.

A method of investigation is used to guide the researcher throughout the study and to help in achieving the research aims and objectives that have been set. In the next section the method of investigation chosen for this study will be briefly discussed and elaborated on in Chapter 4.

1.5   Method of investigation:

The researcher will follow a positivistic research paradigm. This research paradigm is best suited because the researcher should retain a neutral and objective view of reality throughout the duration of the study. The researcher should not reflect his/her opinion regarding the use of social media techniques, strategies and frameworks currently used by companies. The positivistic research paradigm is known as the dominant research paradigm used when studying information systems and information technology in an organization (Gonzalez & Dahanayake, 2007; Orlikowski & Baroudi, 1991:6).

The researcher will make use of surveys as research method. Surveys allow the researcher to have a certain degree of control over the data collected and to manipulate the research design parameters. Data collection will include questionnaires that consist of Likert-scale questions, multiple-choice questions, dichotomous questions (yes or no), as well as open-ended questions. Questionnaires will be used to determine how social media is currently being used in companies. The positivist research paradigm is associated with quantitative analysis which makes use of statistical analysis and mathematical modelling (Oates, 2006:38). The researcher will make use of computer-aided analysis techniques, as well as statistical techniques to analyse the data gathered by means of questionnaires.

Twitter data (tweets) will also be collected and analysed. This data can then be used to perform sentiment analysis and opinion mining. The effect that sentiment analysis and opinion mining can have on customer relationship management will be emphasized by the researcher. Different data- and text mining methods and algorithms will also be investigated, for instance text classification (for example, index term selection, Naïve Bayes Classifier, k-Nearest Neighbour Classifier, decision trees), clustering, etc. After investigating different data- and text mining techniques, the researcher will make use of open-source software where a process can be built to analyse the Twitter data by using a data- or text mining technique. Subsequently, the outline of this study will be discussed.

1.6   Outline of this study

The chapters of this dissertation are divided as follows and a brief description of each chapter is provided. Figure 1.2 presents the layout of this dissertation.

- Chapter 1 – Problem statement: In this chapter, substantiation of the problem statement is discussed and the research aims and objectives are defined. The research methodology, data collection method and data analysis methods are discussed briefly. The aim of this chapter is to serve as an introduction for the study and to give more background regarding the problem that led to the development of the research question.

- Chapter 2 – Literature study: In this chapter, literature regarding social media, the use of different social media platforms in companies and the way social media data can be used with a variety of text mining techniques are discussed. The aim of this chapter is to improve knowledge regarding the value that well managed social media platforms can add to customer relationship management. Social media frameworks used by companies are also investigated.

- Chapter 3 – Mining social media data: In this chapter, emphasis is placed on how customer relationships can be improved by using different data- and text mining techniques to analyze social media data. Different data-mining applications for businesses are discussed. The aim of this chapter is to investigate different data- and text mining techniques and to identify which techniques are suitable when mining social media data for sentiment and opinions of customers.

- Chapter 4 – Research design: The manner in which the research is conducted is discussed in this chapter. The research paradigm, research method, data collection method and the data analyses method will be expanded upon. The aim of this chapter is to indicate the process followed during this study.

- Chapter 5 - Results of questionnaires: The results of the questionnaires are analyzed by using basic statistical measures such as frequency, cumulative frequency, percentage, cumulative percentage, mean comparison, standard deviation, factor analysis and reliability analysis.

- Chapter 6 – Results of Twitter data: In this chapter the process followed, text mining techniques used, data-mining techniques used and results that have been gathered from the techniques will be discussed.

- Chapter 7 – Discussion and conclusion: This chapter is discussion of the results. The purpose of this chapter is to determine whether the aims and objectives set according to the research question of this study were achieved. Limitations and future work pertaining to the study are also presented.

- Bibliography: This section provides a list of authors used as references for this study. The reference style used throughout the dissertation is the Harvard style.

Utilizing social media to the benefit of companies

Chapter 1: Problem statement
1.1 Introduction
1.2 Problem statement and substantiation
1.3 Previous similar studies
1.4 Research aims and objectives
1.5 Method of investigation
1.6 Outline of this study
1.7 Conclusion

Chapter 2 - Literature Study: Social media
2.1 Introduction
2.2 Defining social media
2.3 Different public external social platforms
2.4 The use of social media to build customer relationships
2.5 Conclusion

Chapter 3 - Literature study: Mining social media data
3.1 Introduction
3.2 Data- and text mining in the business environment
3.3 Using social media data for opinion mining and sentiment analysis
3.4 Data-mining techniques
3.5 Conclusion

Chapter 4: Research design
4.1 Introduction
4.2 Research paradigms
4.3 Research method
4.4 Data generation method
4.5 Data analysis
4.6 Conclusion

Chapter 5: Survey results
5.1 Introduction
5.2 Statistical techniques used in this study
5.3 Section A – General information
5.4 Section B – Social media platforms
5.5 Section C – Social media strategy
5.6 Section D – Social media metrics
5.7 Conclusion

Chapter 6: Results of Twitter data
6.1 Introduction
6.2 Using Rapidminer Studio
6.3 Importing data into Rapidminer Studio
6.4 Process built in Rapidminer Studio
6.5 Results obtained from Twitter data
6.6 Conclusion

Chapter 7: Discussion and conclusion
7.1 Introduction
7.2 Discussion regarding survey results
7.3 Discussion regarding results obtained from analysis of Twitter data
7.4 Contribution of this study
7.5 Answering the research question and achieving secondary objectives
7.6 Limitations of this study
7.7 Future work and research
7.8 Conclusion

Figure 1.2 - Outline of this study

1.7   Conclusion

The aim of this chapter is to serve as an introduction to the dissertation and the research. This chapter provides the problem statement (§1.2), namely that companies are faced with challenges regarding the use of social media. Challenges are, for example, choosing a suitable social media platform(s), accepting the risks and limitations of social media and measuring the effectiveness of social media platforms. All these challenges can also influence how a company approaches customer relationship management done through social media platforms. Previous studies do not indicate which techniques, strategies and possible frameworks a company can use when using social media.

The research aims and objectives (§1.4) have been formulated according to the research question identified during the initial phase of this study. The research question, which the researcher will answer at the end of this study, is: Which techniques can be used by a company to ensure that social media is utilized effectively? This study and research will contribute in improving knowledge regarding the value of social media if managed correctly, providing companies with a framework and available techniques that can be used to analyse customers'/clients' feelings and paradigms about the company, and which can then be used to improve customer relationship management.

The researcher will implement a positivistic research paradigm. The positivistic research paradigm was chosen because the characteristics of the study align with characteristics of this research paradigm. A positivistic research paradigm makes use of the scientific methodology, which will allow the researcher to act as an impartial observer in a world that exists independently of humans. The researcher will be able to gather quantitative data and perform quantitative data analysis by using mathematical and statistical modelling and analysis.

This dissertation is divided into seven chapters and the outline of the dissertation was discussed. To be able to understand the use of social media in companies and the problems and challenges that companies are facing when utilizing social media, literature on the subject should be studied. A literature study was done to gain better understanding of the problem stated in this chapter.

CHAPTER 2: LITERATURE STUDY – SOCIAL MEDIA

2.1   Introduction

In this chapter, social media, the use of different social media platforms in companies and the way social media data can be used with a variety of data- and text mining techniques are discussed. In Figure 2.1 an overview of the themes discussed in this chapter is illustrated.   Terms related to social media, different public external social platforms, problems, challenges and boundaries that a company can come across when making use of social media, and how social media can contribute to social capital and customer relationship management are discussed in this chapter

In this chapter an attempt will be made to improve knowledge regarding the value of social media if managed correctly and will also contribute to achieving a secondary objective set by the researcher in Chapter 1. The secondary objective (SO1) is to review literature on the use of social media in companies and by doing so, techniques and frameworks currently used by companies can be identified. The researcher can then determine which techniques can be used effectively by a company to utilize social media to the benefit of the company and to improve the company's knowledge regarding customers/clients.

Utilizing social media to the benefit of companies	
Chapter 1: Problem statement	
Chapter 2 - Literature Study: Social media **2.1 Introduction** **2.2 Defining social media** Terms related to social media for example Web 2.0 and User-Generated Content (UGC) is discussed. **2.3 Different public external social platforms** External social media platforms such as Facebook, Twitter, LinkedIn, YouTube and Google+ are discussed. Different types of media can be shared on these platforms. The problems, challenges and boundaries of using social media are also discussed. **2.4 The use of social media to build and improve customer relationships** In this section social capital and customer relationship management when using social media is discussed. **2.5 Conclusion**	**Chapter 3 - Literature study: Mining social media data**
Chapter 4: Research design	
Chapter 5: Results of questionnaire	**Chapter 6: Results of Twitter data**
Chapter 7: Discussion, interpretation and conclusion	

Figure 2.1 - Overview of Chapter 2

In section 2.2 of this chapter, the term social media is defined, and an overview of the differences between Web 1.0 and Web 2.0 is given. In this section, the categories of Web 2.0 tools, which include blogs and podcasts, social networks, communities, content aggregators, and virtual worlds, are also discussed. In section 2.3, different social media platforms are discussed and compared according to the way that social networking site(s) work and the type of user content generated by the site, as well as how the site(s) can be utilized in a company.

In section 2.4 different problems, challenges and boundaries of social media, which a company might want to consider, are investigated. The term social media fatigue is also defined and recommended solutions for the different problems and challenges are discussed. In section 2.5, a brief description follows regarding social capital and the building of customer relationships when using social media.

2.2   Defining social media

Kaplan and Haenlein (2010:60) claim that there seems to be confusion among managers of companies and academic researchers about the term social media. Kärkkäinen *et al.* (2010) define the term social media as "applications that are fully based on user-generated content and this user-generated content or user activity plays a significant role in increasing the value of the application or the service." Kaplan and Haenlein (2010:61) define the term social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that offer the opportunity to create and exchange User-Generated Content (UGC)."

The term social media can be broken up into two terms that can be discussed separately, namely the terms social and media. The Reader's Digest Oxford Complete Wordfinder (1993:1472) defines the term social as "relating to society or its organization" and indicates that to be social a person is "concerned with the mutual relations of human beings or of classes of human beings." The term media is defined as "the main means of mass communication" (Reader's Digest Oxford Complete Wordfinder, 1993:947).

A definition of social media can be formed from the above-mentioned definitions; Social media is Internet-based applications, which create the opportunity for users to create and share content and activities, which may include different forms of media, such as text, images, videos, etc. To fully understand the term social media, the following two terms

should also be understood: Web 2.0 and User-Generated Content (Dijkmans *et al*, 2015:58; Kaplan & Haenlein, 2010:60). These terms will be discussed in §2.2.1 and §2.2.2.

2.2.1   Web 2.0

Web 2.0 can be defined as technologies that enable users to communicate, create content and share this content with other users by means of social networks, communities and virtual worlds (Kärkkäinen *et al*, 2010). Web sites that include a strong social component and several different concepts and technologies, for example user profiles and friend links can be categorized as Web 2.0 web sites (Cormode & Krishnamurthy, 2008). A new generation of web interfaces enables users to read, share and write content by using the World Wide Web; this is described by the term Web 2.0 (Balasubramaniam, 2009:28).

Web 2.0 is the second phase in the evolution of the Web. Web 2.0 is more dynamic and interactive than Web 1.0, as users can access and contribute to the content of a web site. There are a couple of differences between Web 1.0 and Web 2.0 and it is difficult to categorize some web sites as either Web 1.0 or Web 2.0 (Cormode & Krishnamurthy, 2008). Social web sites, such as Facebook and MySpace are regarded as a prototypical example of Web 2.0. This is because of the social-networking aspects, which are included (Cormode & Krishnamurthy, 2008).

Site features can be used to distinguish between Web 1.0 and Web 2.0. Important features that indicate a Web 2.0 website include (Cormode & Krishnamurthy, 2008):

- The user is a first-class entity and owns a profile page which may include demographic information, such as age, sex, location, marital status, occupation etc.
- Web 2.0 allows the user to form connections, for example a link to another user who is a "friend" or falls in a specific circle/group of the user. A user can also be exposed to a variety of groups, subscriptions or RSS feeds (Really Simple Syndication, also known as Rich Site Summary) of updates from other users.
- Web 2.0 enables the user to post different forms of media/content, for example photos, videos, blogs, comments, rate another user's content, tag content and control privacy and sharing.
- Web 2.0 includes technical features, such as public API's (Application Programming Interface), which allows third-party enhancements and embedding of various rich content types. API's also allow the user to communicate with other users through Instant Messaging systems or internal email systems.

- Technologies used in the development of Web 2.0 web sites include AJAX (autonomous Javascript and XML). The use of XMLHttpRequest allows a web site to update a page without explicit reload actions. These technologies are incorporated because Web 2.0 web sites involve dynamically generated pages. A Web 2.0 web site can be updated while the user is using the site. This must be done without the user realizing that the site has been updated and the site will inform the user that the site has been updated only when the user requests an update during the visit to the site.

Murugesan (2007:35) lists differences between Web 1.0 and Web 2.0. The following differences listed are related to Web 2.0:

- Web design is flexible with creative reuse and updates.
- The user interface provided is rich and responsive.
- Collaborative content creation and modification is facilitated.
- By reusing and combining different applications or by combining data and information from different sources on the Web, new applications can be created.
- Establishes social networks of people with common interests; and
- Helps gather collective intelligence and supports collaboration.

Most Web 2.0 technologies have been introduced within business-to-consumer markets (Lehtimäki *et al,* 2009). Consumers nowadays are more reliant on information, which is received from peers, due to Web 2.0 and the peer-to-peer on-line communication that it offers. Web 2.0 creates the opportunity for businesses to more effectively reach their target audience and build relationships with current and potential customers. Web 2.0 technologies include different Web 2.0 tools that users of social media utilize to generate content. In §2.2.1.1, Web 2.0 tools are divided into five main categories. The five categories include: blogs and podcasts, social networks, communities, content aggregators and virtual worlds.

2.2.1.1.   Web 2.0 tools

Interest from businesses to determine how different Web 2.0 tools can be used for marketing activities has grown. It is evident that Web 2.0 may have significant effects on a business environment. Web 2.0 tools and technologies offer the opportunity for users to select, filter, publish and edit information (Jussila, 2011). Kaplan and Haenlein (2010:61) are of the opinion that there is no systematic way to categorize social media applications.

It must be kept in mind that new social media sites appear on a daily basis and that the classification scheme takes into account forthcoming applications. According to Lehtimäki *et*

al. (2009) and Constantinides and Fountain (2008), there are two types by which social media activity can be measured, namely on-going analytics and campaign-focused metrics. On-going analytics monitors and tracks activity over a specified time and campaign-focused or event analytics has a clear beginning and end.

The five categories of social platforms (Web 2.0) identified by Lehtimäki *et al.* (2009) and Constantinides and Fountain (2008) follow and is a compilation from the mentioned sources:

1. *Blogs and podcasts*: This tool consists out of traditional blogs, vlogs, podcasts and videocasts. The focus of this Web 2.0 tool is on informing people of current events and novelties. Blogs and podcasts is an easy and cheap tool to maintain, but a weakness is that it requires time and constant updating. The number of people who visit blogs and podcasts is usually measured by the number of viewers, comments and downloads.

2. *Social networks*: Social networks focus on content sharing, maintaining relationships and networking. It is easy to set up a profile and businesses may consider social networks as a possible method that can be used for advertising, but the big question that remains is how to persuade users to participate. The participation and exposure of a social media platform are measured by the number of members or fans connected to a profile or by the amount of user-generated content that is generated in a social media community.

3. *Communities*: Communities are categorized into on-line communities, content communities and forums or bulletin boards. Communities ensure intense two-way communication, but require lots of resources to maintain. The measuring method used for this Web 2.0 tool is the number of members and the quantity of user-generated content in the community. Communities in general can be divided into three different types: on-line communities, content communities and forums or bulletin boards, a description of each follows:

   - On-line communities consist of three different types, namely member-initiated, organization-sponsored and third-party established. Member-initiated communities focus on members' mutual interests and interaction. Organization-sponsored communities tend to focus on business transactions, brand building, interaction between a business and customers and co-creation of products. Third-party established communities allow for communication and transactions between buyers and sellers. Table 2.1 depicts four different types of on-line community and indicates the focus area of the community, as well as why consumers and companies choose to participate in an on-line community.

- Content communities consist of content-sharing sites and wiki's. The main focus is content sharing.
- Forums or bulletin boards focuses on discussion of mutual interests.

Table 2.1 - Classification of online communities (Lehtimaki et al., 2009:30)

Type of Online Community	*Focus*	*Why consumers choose to participate*	*Why companies choose to participate*
Member-initiated communities:	The interests and hobbies of the members are mutual.	Consumers participate to bond with other consumers and share information/content. Consumers also participate to emphasize their individuality.	Companies participate because of the opportunity for targeted advertising.
Organization-sponsored communities also known as brand communities:	Organization-sponsored communities are more focused on brand building, business transactions and maintaining consumer relationships.	Consumers participate to share product information and content. Consumers are also attracted to the fact that their identity in the community is emphasized. Consumers also enjoy participating in the product development process.	Companies participate to reach their target audience and get feedback from consumers about a specific product or service. Companies enjoy that the consumer is part of the product development process. It also displays loyalty to the consumers.
Third-party established communities:	Tend to focus on business transactions and the marketplaces are maintained by intermediaries.	Consumers tend to participate seeing that it is a safe environment in which business transactions can be done.	Companies participate because of the opportunity for targeted advertising.

4. *Content aggregators*: The tools used in content aggregators consist of RSS (Rich Site Summary), widgets, bookmarks, tagging services, etc., and mainly focus on categorizing and customization of web content. Content aggregators are also easy to use, but content needs to be interesting enough to be tagged. By measuring the number of subscribers or tags as well as downloads made, the popularity of this Web 2.0. tool can be determined.

5. *Virtual worlds*: Virtual worlds serve as a substitute for the real world. Virtual worlds engage customers effectively and virtual worlds require a great quantity of resources to maintain and to induce users to participate. The measurement method is users' participation activity.

As can be seen in the above discussion, each of these five categories has a measurement method for user activity and this is summarized in Table 2.2.

Table 2.2 - Measuring methods used by different Web 2.0. tools

Type of Web 2.0 tool	Measuring method
Blogs & podcasts:	The number of viewers, comments and downloads.
Social networks:	The number of members or fans in the profile as well as the quantity of user-generated content in the community.
Communities:	The amount of members and the amount of user-generated content in the community.
Content aggregator:	The amount of subscribers or tags as well as downloads.
Virtual worlds:	Users' participation activity.

2.2.2 User-generated content (UGC)

Vast quantities of user-generated content are created on a daily basis through social media platforms and the amount of content grows exponentially. It is critical for companies to have effective management of user-generated data and that this data is utilized correctly (Gundecha & Liu, 2012:3). User-generated content (UGC) currently has no uniform definition, but can be viewed as a main feature of the user participative web (Balasubramaniam, 2009:28; Cormode & Krishnamurthy, 2008; Clever *et al,* 2009).

Kaplan and Haenlein (2010:61) are of the opinion that Web 2.0 represent the ideological and technological foundation and user-generated content is the way by which people tend to make use of social media. Media that includes radio, television and printed publications are not found to be as efficient as the on-line environment and consumers are directing their attention to interactive media (Daugherty *et al.,* 2008:16). The Web, Web 2.0 technologies, and user-generated content offer the opportunity for consumers to tailor their exposure to media according to their specific needs and desires (Daugherty *et al.,* 2008:17).

When consumers' needs and desires are considered, more focus is placed on consumer-centric media rather than on publisher-centric media as was the case in previous years. An external motivation, which influences the consumer to create content does not exist. By way of explanation, the business will not influence the consumer to create content. The consumer's behaviour would also have an influence on the content that is being created (Daugherty *et al.,* 2008:17).

The process of creating content is not only rewarding for the user. Posting content on a social media platform creates a sense of belonging for some users, but can also offer the opportunity to gather data (Krumm *et al.,* 2008:10). By analyzing the content posted by a user/consumer, a business can gain data and information, which can lead to improving decision making regarding future marketing campaigns, brand awareness and customer relationship management.

Balasubramaniam (2009:30) identifies four major drivers that have an impact on the generation of user-generated content and the use of social media platforms:

1. *Technological drivers:* As broadband and high-speed Internet improve, it allows users to upload and download a massive quantity of data, for example large videos or pictures. New technology advances in devices, such as digital cameras, digital video recorders, smartphones, etc. that is being developed, also allows users to create and upload more content, which can then be uploaded.

2. *Social drivers:* The number of people using social media platforms and the Internet increases daily as people who were born in the Information Technology era are more familiar with distributing content on the World Wide Web. Companies and organizations are also using social media platforms to start new trends.

3. *Economic drivers:* Pressure is placed on companies to use a user-generated content platform to market a brand or launch marketing campaigns that are viral. Costs and availability of tools should be considered by companies.

4. *Legal and institutional drivers:* These drivers are a key factor in the development of user-generated content. Copyright is granted to users to be able to use their content. This is done by using end-user licensing agreements.

User-generated content can be divided into different forms: media web sites (e.g. YouTube), chat interfaces (e.g. Facebook, LinkedIn and Twitter), platforms that allow users to share personal information (e.g. Flickr), ecommerce platforms (e.g. Ebay and amazon.com), and blogs that are used to represent news and inform people (Balasubramaniam, 2009:30). Media web sites, chat interfaces, and platforms that allow sharing of content are known as external public social media platforms.

2.3   Different public external social platforms

It is important to distinguish between the different types of social platform in general. According to a survey done by Social Media Today and Leader Networks (2013:9), a company can choose from three different types of social platform. Firstly, the company can choose to make use of public external platforms, for example Facebook, Twitter and on-line support forums. A second option is to make use of internal social technologies, which include the development of an Intranet and consists of on-line communities. The last option that a company can decide to use is a private external platform that includes on-line communities for customers or business partners (Social Media Today & Leader Networks, 2013:9). Internal social technologies and private external platforms are defined as closed networks.

In this section, the top five public external social platforms will be discussed, according to Social Media Today and Leader Networks (2013:11), Facebook, Twitter, LinkedIn, YouTube, as well as Google+. It should be kept in mind that other public external social platforms also exist, for example Pinterest, Instagram, Flickr, Tumblr, Slide Share, Foursquare, Path, etc. The use of social networks in the business environment and by companies will be discussed. Social networks can be used for a variety of purposes in the business environment, namely as marketing tool, for market research, as communication tool, as a tool, which could lead to increasing a company's web site traffic and sales, as a customer service, for identity development, etc. (Geho, 2010).

2.3.1   Facebook

According to Kwok and Yu (2013:86), Facebook has received considerable attention from many disciplines.   Social networks provide social and emotional support, as well as information resources to users (Joinson, 2008:1027; Kwok & Yu, 2013:86). Facebook allows users to create a profile space, facilities to upload different types of media, various forms of messaging and the opportunity to connect with other users (Joinson, 2008:1027). Facebook does not only offer communication opportunities for individuals, but also for companies seeking to build relationships within the company and with customers and business partners.

2.3.1.1   The case of Facebook

Facebook was launched in 2004 by then-students Mark Zuckerberg, Dustin Moskovitz, Chris Hughes and Edouardo Savarin (Goodfellow, 2012:6). Facebook has become one of the most popular social networking sites that grows by a number of users on a daily basis (Goodfellow, 2012:6; Ross *et al.,* 2009:579). Facebook was developed for the purpose of allowing students of Harvard University to create and maintain relationships formed during their years at university. The volume of daily status updates, photos and other user-generated content of an individual user doubles every year (Bradshaw, 2011). A quarter of all South Africans are using Facebook (Van Zyl, 2015; World Wide Worx, 2016). South African Facebook users mostly access this social media platform via mobile devices (Van Zyl, 2015; World Wide Worx, 2016).

Facebook allows users to follow the actions, beliefs and interests of other users. An important part in a social networking site is the user's profile (Lampe *et al.,* 2007:436). A Facebook users' emotional connection to a page may increase the page's ease of use and usefulness. A user's personality characteristics might also play a part in the on-line

communication experience of a user (Ross *et al.,* 2009:59). A clear distinction is drawn between the use of social networks, such as Facebook, as a 'social searching' site and for 'social browsing' (Lampe *et al.,* 2007:435). 'Social searching' refers to a user searching for information about off-line contacts and 'social browsing' refers to a user making use of a social networking site to develop new connections.

2.3.1.2   Facebook in the business environment

Only a few of the studies done have focused on how Facebook can be used as an effective business-to-consumer relationship building tool (Kwok & Yu, 2013:86; Waters *et al.*, 2009:102). Up to 2009, no guidelines existed, which a company could make use of to manage its social media presence (Waters *et al.,* 2009:102). By using Facebook and other social media applications, companies build a relationship and interact with the public, as well as stakeholders. Social networks and internet communication, for example the instant messaging function of Facebook, do not have to create distance between a company and the public (Kent & Taylor, 1998:323).

There are different uses of Facebook in the workplace, such as reconnecting, maintaining awareness and keeping in touch, as well as building social capital (Skeels & Grudin, 2009:99). The use of Facebook and similar social networking sites are used as a communication tool for businesses of all types and sizes (Crews & Stitt-Gohdes, 2012:2). Seeing that Facebook was originally created to allow students to connect, this platform allows young employees to reconnect with former classmates or with colleagues from previous jobs.

Status updates may help employers and users remain up to date with trends in their field for example, posts from professional colleagues about new technology and changes in an industry over time (Skeels & Grudin, 2009:99). Users are of the opinion that stronger relationships are built and bonding occurs because of the exchange of personal information (Skeels & Grudin, 2009:99).

Posts made by consumers on a business's Facebook brand page can reveal the following; perception of the brand, acceptance of a new product, most favoured products and features, required products and features, problems, locations with great sales volumes. This can create the opportunity to generate new ideas, and identify competitors (Cvijikj & Michahelles, 2011:6). Research carried out indicates that product, sales and brand are the three most discussed topics on a business's Facebook brand page. The most common

intentions for participation by consumers are requests and suggestions, expressing affect and sharing.

2.3.2 Twitter

Microblogging has become a very popular form of communication among Internet users (Pak & Paroubek, 2010:1320). Users tend to include their opinions regarding certain topics and issues in messages on social networking sites, such as Twitter, Tumblr, Facebook, etc. (Pak & Paroubek, 2010:1320). Users also post messages regarding products that they have bought, services they have used, customer services provided by a certain company, as well as political and religious views.

2.3.2.1   The case of Twitter

Twitter is a popular microblogging, on-line social network, used by millions of people connected around the world and launched in 2006 (Kwak *et al.,* 2010:591; Sakaki *et al.,* 2010:851). By August 2015, 7.4 million South Africans were using Twitter (Van Zyl, 2015; World Wide Worx, 2016). Microblogging includes the sending of brief text updates or micro-media (Sakaki *et al.,* 2010:851). Each message, also known as a *tweet,* is limited to 140 characters and can reference one or more places, which can be mapped to real-world locations (Johnson & Yang, 2009). The character limit of statuses makes Twitter different from other social networking sites (Geho *et al.,* 2010).

A user profile includes the full name, the location, a short biography, as well as the number of tweets that the user had made thus far (Kwak *et al.,* 2010:592). Users can select from which other Twitter users they wish to receive updates. This is known as "following" a user and received updates are viewed from the same interfaces (Johnson & Yang, 2009).

The people who follow the specific user are listed, as well as the people who the user is following. Twitter follows trending topics by tracking phrases, words and hashtags that are most often used in users' tweets and posts (Kwak *et al.,* 2010:592). A hashtag is a convention, which is used among Twitter users. The '#' character is used as a prefix to a word, which will allow other users to follow a thread of discussion.

On the right side of the sidebar of each user's homepage, the top ten trending topics of the moment are displayed by default. This option can be set by the user. It should be kept in mind that the trending topics are not necessarily grouped by Twitter (Kwak *et al.,* 2010:592). Twitter also supports one-to-one, one-to-many and many-to-many communications (Zhang

et al., 2011:162). Other examples, similar to Twitter, include Tumblr, Plurk, Emote.in, Squeelr, Jaiku, etc. Each of these has unique characteristics by which they can be distinguished.

2.3.2.2   Twitter in the business environment

Twitter is considered to be an easy access point for businesses that decide to utilize social media for the benefit of their company. The use of Twitter as a marketing tool by a business or company is simple, not expensive and the learning curve for Twitter is short (Geho *et al.,* 2010). Twitter can be used by companies to create and maintain a network with professionals and to stay informed about and communicate the latest industry news (Geho, 2010). Twitter can also be used as a market research tool to examine whether the products or services being placed in the market are appropriate to the consumers' needs.

Zhang *et al.* (2012:161) examine the use of Twitter as an electronic word-of-mouth (eWOM) communication form and are of the opinion that eWOM messages can be used as a marketing tool to create more exposure for a company's products and services. If a company is using Twitter as marketing tool or promotional opportunity, professional marketing strategies are seldom used (Geho, 2010). There has been a power shift between companies and customers due to today's technology-driven business market and this has resulted in companies considering a new marketing strategy and paradigm (Geho, 2010).

Twitter, including other social networking sites, is considered as a free advertising platform (Geho, 2010). Based on the number of followers that a company has generated, an opt-in charge may be charged when users want to access specific information, which is tied to Twitter. These charges and search features of Twitter may lead to Twitter generating a revenue stream. As Twitter grows on a daily basis, there is a possibility that this social networking site might charge a fee for commercial users in the nearby future (Geho, 2010). This is not necessarily a negative aspect because Twitter-user fees could provide a great opportunity and environment for its business advertisers (Geho, 2010).

2.3.3 LinkedIn

LinkedIn can be described as a tool which allows a user/company to identify and contact appropriate candidates to fill new or vacant job positions (Stelzner, 2014; Thew, 2008:87). It is believed that LinkedIn is an invaluable business networking tool. LinkedIn can be used for various purposes, whether personal or business, such as a recruitment tool or as a source of information regarding industry news and developments.

2.3.3.1　The case of LinkedIn

According to Skeels and Grudin (2009:96), LinkedIn was developed as a professional social networking site and users can be divided into three groups: current or recent students, young professionals, as well as older professionals. According to studies done by Skeels and Grudin, there is a clear indication that the use of LinkedIn is more popular under young professionals than other users (Skeels & Grudin, 2009:97). This social media network focuses on the distribution of professional information and on encouraging users to construct abridged curriculum vitae, which will allow other professionals to connect with a user according to experience and qualifications.

Different from other social media networking sites, such as Facebook, LinkedIn users' profiles are kept strictly professional, and personal information, such as hobbies, favourite books, movies, music, etc. are not deemed important factors (Skeels & Grudin, 2009:97). The visitors/users of LinkedIn do not frequently visit their site or those of friends. A user's LinkedIn page is relatively static apart from any new connections that have been made (Skeels & Grudin, 2009:97). However, users can make a recommendation to other members and can explore the direct connections from their connections. Paying members can use the opportunity to search for LinkedIn users who comply with certain occupational characteristics. This is an option that is very convenient for recruiters or consultants (Skeels & Grudin, 2009:97).

2.3.3.2　LinkedIn in the business environment

According to Thew (2008:88), LinkedIn creates a great opportunity for building a network and making connections. LinkedIn enables users to continuously be aware of other users' activities. By reviewing connections, a user can see which connections or users have changed jobs or who may be unemployed and companies can use LinkedIn as a recruitment tool (Skeels & Grudin, 2009:98). Users can indicated whether they are interested in new career opportunities, finding new business contacts or just keeping in touch with contacts by using a drop-down menu with these mentioned categories (Thew, 2008:88).

An example of where LinkedIn is suited, is if a person meets a possible business connection at a social event, such as a conference and there is an exchange of business cards. The person wants to stay in touch with the possible business connection, but once a business card is filed, the card may well remain untouched. The person may feel that sending an e-mail is too imposing because a reply is expected and these types of e-mail are rarely sent.

By using LinkedIn, an invitation can be sent, which will allow the possible business connection to accept without feeling pressured (Skeels & Grudin, 2009:98).

When a user invites another user on LinkedIn, the user is also inviting the person to join his or her personal network of contacts (Thew, 2008:88). Building a network is described as being straightforward, and a variety of methods, which can be used, exists. One method, which can be used to build a network is by sending out invitations in batches (Thew, 2008:88).

LinkedIn offers the user the option to download a toolbar which can be used to upload contacts from Outlook mailboxes, to select the contacts to which the user wants to send an invitation to and to send the group of invitations directly to the specified e-mail addresses of the selected contacts. The growth of the network will depend on the response of the individuals that the user has invited, but it has been found that the response is usually positive (Thew, 2008:88). Each user who accepts the invitation can then repeat the process by using his/her own contacts and unknowingly the network grows.

LinkedIn offers the user different tools, which can be used to search for new connections by name, company, location, type of contact, etc. Another search option that can be used to find new connections is the advanced search option, which allows the user to improve the level of precision of the search (Thew, 2008:88). The results of a search will be ranked in order of proximity to the user and the user's network. The users linked to this new connection will be displayed. LinkedIn also offers the option to form a group. This application is formal and there is an acceptance process through which the newly formed group will have to go through. A group may consist of a network of alumni, a particular company's employees or a professional organization or interest group, which is usually convenient for market research.

2.3.4   YouTube

YouTube is a social networking platform that focuses on distributing video content. It is a simple to use social networking site, which businesses make use of to market a product or to develop and build a brand and image (Haridakis & Hanson, 2009:317). The distribution of videos enables a business to not only share textual content with customers, but also allows the customer to view content that carries information regarding a company in a visual format.

2.3.4.1   The case of YouTube

According to Gill *et al.* (2007:16), YouTube was founded in 2005 as a web site, which allowed users to share video content. Haridakis and Hanson (2009:317) describe YouTube as an on-line form of communication that allows users to be both a consumer and transmitter of content. YouTube also illustrates the speed by which content can be distributed through social networking innovations. The content on YouTube is either published by traditional mass media or uploaded by YouTube users (Haridakis & Hanson, 2009:318).

A user of YouTube can upload a video clip in most standard video formats, for example WMV, MPEG and AVI, which is then converted to low resolution Flash format (swf) when it is uploaded (Gill *et al.,* 2007:16; Paolillo, 2008). YouTube makes use of Adobe's Flash Video (FLV) format for delivery (Gill *et al.,* 2007:16). A user must create a user account, also known as a channel before he/she can start uploading videos (Paolillo, 2008). An account includes a profile page, which displays a list of the user's uploaded videos, as well as other channels, which the user is following. As YouTube grew, additional features were added to accommodate users' social networking demands.

These features allow users to tag a video by using keywords and phrases, which describe the content of the video. The sharing of user-generated content allows YouTube to be classified as part of Web 2.0. YouTube also provides the option for users to share their opinions about the content by leaving on-line comments, rating a video and sharing content by e-mailing hyperlinks to connections (Haridakis & Hanson, 2009:318). Users of social media, such as YouTube will visit the site to satisfy underlying needs of interests and curiosity (Haridakis & Hanson, 2009:318).

2.3.4.2   YouTube in the business environment

Companies are encouraged to embrace the use of social media and not to fear that control is being lost regarding the company's brand and identity (Waters & Jones, 2011:252). Text-based conversations are a great way to develop and build an organization's reputation and it can also help with the development of the company's brand, but when text is combined with a visual element, the identity of a company is enhanced (Waters & Jones, 2011:252). If a company uses videos to build an image and share the video not only on YouTube, but also on various social networking sites, conversations will be started by consumers and this will help build customer relationships and increase a company's reputational yield.

YouTube offers the opportunity to bridge the gap between search, discovery, content, video, as well as social (Spiegel, 2014). YouTube offers valuable text content and also collects a wide variety of data points (e.g. views, likes and votes). For a company to have a successful product and marketing campaign, the company must understand the needs of the consumer. On YouTube the required data can be collected, cleansed and filtered. An analysis can be run on the raw data, which will lead to the calculation of frequencies and occurrences (Spiegel, 2014). The insights and findings can then be extracted.

2.3.5 Google+

After Google introduced Google+ as a new social networking site, a significant growth in its size has been reported (Gonzalez *et al.,* 2012). Similar to other social networking sites Google+ offers similar features to other social networking sites and includes a couple of new features as well. Google+ pages allow companies and businesses to distribute more information about the product, brand or service, as well as the company itself (Google, 2015).

2.3.5.1   The case of Google+

According to Kairam *et al.* (2012:1066), Google+ is a social networking service, which was launched in 2011 by Google. This social networking site is regarded as a new generation in social networking. Google+ introduced *circles,* which is similar to groups or lists formed in Facebook and Twitter. By making use of different circles, the user is able to organize his or her contacts (Magno *et al.,* 2012:159*;* Kairam *et al.,* 2012:1066). Each time a user chooses to share content, the user is able to decide on the audience with which he or she wishes to share the content. A user can choose to share the content publicly or selectively with one or more of the user's circles (Kairam *et al.,* 2012:1066). By selecting an extended circle option, the user can also share content with friends of his or her connections.

For example, a user may have the following circles, "family", "colleagues", "alumni", etc. When a user adds another person to a circle, the user will receive updates from this new connection (Magno *et al.,* 2012:160). A user can also set the visibility of content being shared and which of the user's circles will receive the shared content. Other new features that Google includes are "hangouts", which allows users to create a video-chatting session and to invite up to nine people from their circles to share in this experience, photo albums, which allows a user to upload, share and organize photos and a messaging service (Magno *et al.,* 2012:159).

2.3.5.2   Google+ in the business environment

Many companies are uncertain about the use of Google+ (Hines, n.d.). One of the reasons why companies feel alarmed when using Google+ as a branding and promotional tool is because Google previously attempted social networking with sites, such as Orkut, Google Buzz and Google Wave (Hines, n.d.). Google+ pages offer businesses and companies the opportunity to set up a profile specifically for their business and brand. Google+ pages offers two types of page, namely local business pages and brand, organization, or artist pages.

A local Google+ page includes features that allow a customer to easily connect with the businesses' physical location and to provide accurate information about a business/company (Google, 2015). Brand, organization, or artist pages do not include a physical location, but will allow information to be sent out to followers, fans and customers on Google (Google, 2015). A clear policy states that brands and businesses are not entitled to running "contests, sweepstakes, offers, coupons or any other promotions" (McCarra, 2011). A company that wishes to promote their brand in such a way, should provide a link to an external web site or to an alternative social network (McCarra, 2011). Users will follow a company's Google+ page to stay informed regarding the brand and product or service that the company is offering.

2.3.6   Comparison between the different social media platforms

Social media platforms each serve different purposes and meet different needs of users (Gundecha & Liu, 2012:4). Facebook and similar sites are becoming a communication tool which is used by businesses (Crews & Stitt-Gohdes, 2012:2). Businesses have the view that newly hired employees are already using Facebook and other social networks as a communication tool and it would be easy for these employees to add new colleagues to different social networks. It must be kept in mind that users do not only use social networks as a form of communication, but also to keep in touch, to re-acquire lost contacts, to make new contacts, etc. (Joinson, 2008:1029). This feature is similar for each social media platform.

In this section, a comparison between the five different external social media platforms, discussed in previous sections (§2.3.1 to §2.3.5), is done according to the different key features of each platform, as well as other features. Table 2.3 on page 28 is a summarized comparison of the five different external social media platforms. The content or type of media that can be distributed on a social media platform differs from platform to platform.

For example, text can be used in various ways to post a status or comment on another user's status or tweet. Each platform has its own unique method by which user-generated content can be created.

Facebook offers an enormous network that is combined with flexibility and simplicity of use, which makes it a good fit for most companies (Wiegand, 2015). A company can benefit from using Twitter. Twitter offers a steady flow of news flashes, as well as links, which are shared by followers. Twitter also offers the opportunity for companies to target new customers and to increase the visibility of a brand (Wiegand, 2015). LinkedIn is also popular among freelancers as a networking tool to identify potential clients. A company can easily use more than one external social media platform, but none of these platforms will help the company to gain customers if time and effort is not invested (Wiegand, 2015).

A company that has a Google+ presence enjoys increasingly favourable rankings when the company's name is searched when Google is used as a search engine. Even though the platform's size is questioned at times, this platform allows a company to have a tailored conversation with customers (Wiegand, 2015). Each platform also attracts certain types of users. Users of social media sites range from individual contributors to senior management and the use of a social media site can also be correlated with age (Skeels & Grudin, 2009:102).

Table 2.3 - Comparison between five external social media platforms

	Facebook	*Twitter*	*LinkedIn*	*YouTube*	*Google+*
Description	Facebook is a social networking site that allows users to keep in touch and up to date with connected friends on the network.	Twitter is a popular microblogging, online social network which is used by millions of people connected around the world.	LinkedIn is a professional social networking site that current or recent students, as well as young and older professionals join to help build networks for business use.	YouTube is a social network that allow users to share video content.	Google+ is a social networking service that introduced *circles* which is similar to groups or lists formed in Facebook and Twitter.
Type of content/media that can be shared	- Text for example: status updates, messages to other users, etc.; - images; - photos; - Videos; and - hyperlinks to other websites.	- Messages known as *tweets* that consist out of 140 characters (A message can also include a reference to another user or product, as well as a hyperlink). - Photos can be shared in real-time with other users and the user can tag connections/friends.	- Content can be categorized for example posts that have been made. - Content from SlideShare can be embedded. - Other content, such as PowerPoint presentations, Keynote presentations, PDF, white papers, infographics and embedded videos from YouTube can also be shared on LinkedIn.	- The main content that is shared on YouTube is video. - From this content other forms of media can also be shared, for instance if a user makes a comment on a video it is text. - Users can also share images.	- Text for example a status or idea, - images, - videos, and - any kind of URL.
Process for becoming a member of this social network site	A person can sign up by providing their first name, last name, email address or mobile number a password, specify their birthday and sex. When a company wants to create a page there are six different options that can be chosen from: - local business or place; - company, organization or institution; - brand or product; - artist, band or public figure; - entertainment; and - cause or community.	A person can sign up by providing their full name, phone or email and username.	If a person chooses to join LinkedIn, the first name, last name, email address and a password which is six or more characters long should be provided.	If a user has a Google account they can upload videos by using this account. A user/company can also create their own channel by signing in to YouTube and filling out the necessary details to create a new channel.	If a user already has a Google account it is easy to join Google+. If a person does not have a Google account the user can type in Google+'s web site and will then be asked to provide their first and last name, choose a username which will be their email address and provide a password. The user will also be asked to enter their birthdate, gender, current email address and their country.
Process to connect with other users	A user can search for the person that he/she wants to send a friend request to. If the person searched for is found a friend request can be send in one mouse click.	If a user wants to connect ("follow") another user on Twitter the user can click on a user name or navigate to the user's profile and then click on the *Follow* button. For a business or company making use of Twitter it is important to create a campaign which will help accelerate the amount of followers and once these followers have been acquired it	A user can send out a connection request to another user. The user sending the request should consider personalizing the request by adding actual content, stating the reason why they would like to connect with the other user, as well as providing a reference of another user. In business it is often that a user would like to introduce two or more of his/her connections with other users. This can be done by sharing	A user can simply follow a company, artist, brand, etc. page by subscribing to the channel.	If a user already has a Google account, the process to connect with other users is simple and a user can accept a person into his/her *circle* by just specify in which one of the *circles* the connection should be categorized. If the user wants to connect with a person who is not a Google account holder, the user can send a Google account invitation which will allow the

	Facebook	*Twitter*	*LinkedIn*	*YouTube*	*Google+*
		is important to continuously engage with them over time.	the profile of the connections that is being introduced.		other user to sing up for a Google account.
Key features	- Facebook offers a subscribe feature which allows a user to follow people without necessary becoming their friend. A person can thus post content on a company's official page without the user being known. - Facebook places an emphasis on visuals for example a user's cover photo to photos placed on their timeline. - Nonprofit companies are using Facebook to help with collecting funding. A user can donate to a nonprofit company/organization simply by clicking on the "Donate Now" button and then filling in certain details. - The newsfeed of Facebook also allows a user to reach content that include promotions, advertisements, etc.	- Requires a short message of 140 characters or less. - Allows users to integrate with other social networks and other web sites. - Hashtags allow users to gain access to other content and to easily connect. - User is in more control of how Tweets can be viewed. - Twitter highlights the most engaging Tweets of the day.	The following five features include benefits if used correctly: - *Headline:* A headline has the same function as the headline of and advertisement. A headline can inform people what the user can do for them and how they can benefit if connecting with this user. - *Groups:* LinkedIn provides the opportunity to create and be a part of a group(s) and by being part of a group it can help a company to amplify their message, build and nurture their network, and provide a source of information regarding the industry and new trends. - *Multi-media:* Multi-media is a great way of building a brand and attracts another user's attention promptly. - *Endorsements:* LinkedIn provides a user to endorse themselves and other uses with skills, the top ten skills for which the user have been endorsed are then displayed. - *Headshot:* By headshot there is meant a profile picture, this will add credibility to the user's profile.	- Organizational channel allows the user to create libraries of videos. - Themed "playlists" are also created. - YouTube is seen as easy to use. - Background analytics can identify who viewers are and what type of videos they are engaging with. - YouTube creates the opportunity for companies to develop their brand and also includes fund raising opportunities. - Also provides the opportunity for businesses to link their web site to a featured video.	- Google+ allows a user to organize other connections into *circles.* - New updates are real-time stream. After a user logs in, new comments will appear continuously as other users comment on the posts. - Another feature of Google+ is Hangouts. This feature allows users to video chat with up to ten people at a time. YouTube videos can be shared with other users during the chat. - Google+ also allows the user to post photos which can be integrated with Google's Picasaweb albums. - Similar to other social networks Google+ allows users to share content, tag other users and create/edit their profile.
Purposes that a company can use the platform for	- Online advertising; - building brand awareness; - highlighting specified products and services; - receiving customer feedback; and - increasing traffic to the company's web site.	- Increasing awareness; - changing perception; - reaching new audiences; and - receiving customer feedback.	- Increasing traffic to the company's web site; - promoting the company on a professional level; and - increasing awareness.	- Online advertising; - building brand awareness; - highlighting specified products and services; - receiving customer feedback; and - increasing traffic to the company's web site.	- Linking authorship of online content (like blogs) with the company's Google+ page. - Linking the company's website with your Google+ page - Building a reputation and network. - Posting regular informative content. - Google Places and Google+ Local is ideal for business use.

2.3.7 General problems and limitations of using social media

Social media may include many challenges and opportunities for companies (Kaplan & Haenlein, 2010:59). It might even be said that companies' business performance and social media may have a love-hate relationship. In this section, the general problems and challenges of social media, problems, such as privacy, violation of user's data protection rights, identity fraud, etc. are discussed. The boundaries that a company should keep in mind when using social media networks are also discussed. A company should evaluate resources that are available before developing a social media marketing strategy (Wiegand, 2015).

According to Wiegand (2015), social media can be used by companies as a marketing tool, but it is not risk free. A company can possibly be exposed to negative publicity because followers are free to post comments on social media platforms. If a company's social media expert reacts immediately to negative comments, the damage can be minimized. Another threat to companies making use of social media is the possibility of being hacked. A company's page or feed can be attacked and false information can then be posted. The correct security precautions should be in place. If a social media platform is not attended to by the company a negative impression can develop (Wiegand, 2015).

The following are general problems and limitations that a company can encounter when using social networks. This section is compiled from two different sources, namely Merril *et al.* (2011) and Rideout (2014:133):

- *Reputation damage:* A company should carefully consider updates and the material uploaded on social network feeds because it reaches followers immediately. Companies place their trust in employees who are tasked with generating user-generated content for social networks. The wrong content, which can be placed unintentionally, can damage a company's reputation. A company should also realize that the amount of time spent on monitoring content should not be neglected. Companies should also train those employees who are responsible for user-generated content. Even if an employee praises the company where he/she is employed, there can be consequences, which can lead to the reputational damage of the company and/or employer.
- *Legal issues:* Companies should consider the legal risks that are associated with using social media. The company's management should consult the inside and outside counsel that is familiar with information technology law. Common situations which may call for legal action include employment risks, security risks, intellectual property and media risks, defamation risks and privacy risks.

- *Ownership of user-generated content:* Junior staff or non-employees are usually appointed and responsible for creating user-generated content. A disengaged employee and/or human-error can easily cause lasting reputation damage to a company. A larger organization should appoint a manager who is in charge of social media output. This manager will then carry the responsibility of approving any user-generated content created before it is published on a social network site. This will reduce the problem of unsuitable content or having sensitive commercial information published.

- *Reduced productivity:* Some companies restrict employees from accessing social networks at work because it has an effect on their productivity. Rideout (2014:138) points out that this argument can lead to a debate because some employers feel that if an employee takes a break, it helps the employee refocus on tasks that still need to be done.

Companies have taken the following approaches to minimize reputational, legal and operational risks (Merril *et al.,* 2011:8). A number of approaches include:

- No employee has access to any social media platforms.
- Only designated individuals will have access to social media platforms and this will be purely for work purposes.
- Employees may have access only to specific sites, such as LinkedIn to further the company's business objectives.
- Every employee has total access at work. This entails that employees should be instructed regarding the comments and content that they generate and publish. An employee will only have a certain amount of time that he/she can spend on social media platforms per work day.

Another issue that may affect the effective use of social media by a company is social media fatigue. A question, which may come to mind for companies when deciding to make use of social media, is whether social media has a life expectancy. Social media fatigue can be defined as "social media users' tendency to back away from social media usage when they become overwhelmed with too many sites, too many pieces of content, too many friends and contacts and too much time spent keeping up with these connections" (Technopedia, 2011).

Factors, such as concerns about privacy and boredom among social media users may also lead to social media fatigue (Bright *et al.,* 2014:149). Social media fatigue follows the trend just as mail, phone, email, web and other technologies have followed (Wang, 2012). Users

will start pulling away from social media platforms when advertisers and companies abuse a social media channel, especially when these users are spammed with unwanted overflow of irrelevant offers. Novas (2015) mentions the term "Facebook depression", which is depression that develops when users, especially teens, spend time on social networks and start exhibiting the warning signs of depression, this is due to the intensity of the on-line world.

Social media fatigue may be of concern for companies that are considering making use of social media (Picardo, 2013). Users who are taking a break from social media still interact by using traditional communication. Incentives are used to help fight the battle against social media fatigue and a company can ensure that the marketing strategy being used is diverse (Picardo, 2013).

To summarize, social media may include many challenges and opportunities for companies. Social media can be used by companies as a marketing tool, but it is not risk free. Possible risks may include negative publicity, creating an opportunity to be hacked, reputation damage, legal issues, ownership of self-generated content and reduced productivity. The issue of social media fatigue should also be considered by companies. To help reduce risks and social media fatigue, a company should be able to use techniques that can ensure the company that a social media platform is being utilized optimally. This statement can be linked to the research question of this study. Next, the use of social media to improve and build customer relationships is discussed.

2.4   The use of social media to build and improve customer relationships

In this section, the term social capital and how customer relationship management can contribute to building social capital of a company are briefly discussed. The difference between customer relationship management and social customer relationship management is described and the different technologies that are associated with each are identified. For customer relationship management to be successful, the company should know and understand its customers and the market (Rygielski *et al.,* 2002:491).

2.4.1   Social capital

Social capital refers to resources that are accumulated through relationships between people and is considered to be a term that has a variety of definitions in multiple fields (Ellison *et al.,* 2007:1145). Building social capital can have a number of positive outcomes, whether for economic or noneconomic reasons (Coleman, 1988:100). Social capital consists

of a number of entities, which have some aspect of social structures and can facilitate certain actions of users (Coleman, 1988:98).

Social capital's sources lie in the social structure within which the user or actor is located (Adler, 2002:18). There are different types of dimension to social structure that are rooted in different types of relation. Adler (2002:18) names three different relations, namely market relations, hierarchical relations and social relations. Social capital can thus be defined as follows: a resource for a company, which includes a variety of social structures and contains actors who contribute to improving the resource.

Ellison *et al.* (2007:1143) describe that social networks can be used in different contexts, such as work-related, romantic relationship initiation, to connect users who share similar interests, etc. Social networks are continuously changing as users of a social network form new relationships and abandon others (Ellison *et al.,* 2007:1148). A company's social capital may be influenced by changes in social networks. Users may use a social networking site to interact with people that they already know off-line.

According to Ellison *et al.* (2007:1145), a distinction can be made between on-line and off-line social capital. Off-line social capital refers to a person that has an off-line connection with another person(s) and then uses this off-line connection and a social media platform to develop an on-line connection with this person. By developing this on-line connection, the person builds his/her on-line social capital.

2.4.2   Conventional customer relationship management (CRM) vs. Social customer relationship management

According to Reinhartz *et al.* (2004), defining customer relationship management (CRM) is a challenge because any definition is contingent to the level of customer relationship management being practised in a company. Customer relationship management is defined as "a management philosophy according to which a company's goals can be best achieved through identification and satisfaction of the customers' stated and unstated needs and wants" (BusinessDictionary, 2015). Parvatiyar and Sheth (2001:5) define customer relationship management as a comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer.

Rygielski *et al.* (2002:491) are of the opinion that customer relationship management can be defined by a framework consisting of four elements: know, target, sell and service. Customer relationship management is an often-studied concept, especially when it comes to business-to-business marketing research, but has no generally accepted definition (Trainor, 2013:318). Customer relationship management is more often used in business-to-consumer organizations and can therefore be compared to business-to-business organizations. For the purpose of this study, customer relationship management will be defined as a customer-centric strategy for building, managing and strengthening long-lasting customer relationships.

Customer relationship management can be viewed as both a strategy and a method, which can make use of information technology to increase customer value and support marketing activities (Trainor, 2013:319). Customer relationship management technology outcomes depend on the development of critical information processes, which result from a business's technology implementation and the business's strategic orientation (Jayachandran *et al.,* 2005:178). Customer-centric business processes and human skills should be integrated with customer-relationship management technology; this will ensure that a business has an advantage over other businesses in the same industry. Customer relationship management involves detailed customer intelligence, which makes it possible to identify the most profitable customers (Rygielski *et al.,* 2002:492).

Most of the studies done to date examine how customer relationship management relates to business performance, but do not examine the role of social media technology implementation or usage (Trainor, 2013:319). In Table 2.4, a comparison of conventional customer relationship management technologies with social customer relationship technologies is presented. Social customer relationship management is not a replacement for conventional customer relationship management. Social customer relationship management can be regarded as an extension of conventional customer relationship management, because it adds social functions, processes and capabilities, which ensure interaction between the consumer and the business, as well as the interaction between the consumer and his/her peers (Greenberg, 2010:414).

Table 2.4 - Comparison between conventional CRM technologies and social CRM technologies (Trainor, 2013:320)

Focus Area	Conventional CRM Technologies	Social CRM Technologies
Customer Relationship	A one-to-one relationship between the consumer and the business exists.	A network of many-to-many relationships exist between the consumer and the business as well as the consumer and peers.

Focus Area	Conventional CRM Technologies	Social CRM Technologies
Customer Communications	Customer communication consists of a one-way monologue.	Communication with customers consists of interactive dialogues.
Technology Use	Technology is deployed to increase employee efficiency/effectiveness. Use of technology is used to focus more internally on marketing/sales.	Technology used to support outside-in processes that connect a business to an external environment and is more consumer-oriented.
Organizational Learning	Information of consumers is viewed as input for decision support systems and internal responses.	Knowledge is co-created with the help of consumers; the arrangement is that there is interaction between the consumer and the business, which contributes to innovation and learning.
Value Creation	Consumers who will contribute to the business are identified (business-centric focused).	Concentrates on dual creation of value: "consumer co-created experiences, collaborative service and support, innovation co-creation" (Tanner, 2013:320).
Data Sources	Information is delivered from the consumer to the business.	Information is mostly created through interactions with and between consumers.
Information Processes	Consumer information is captured and made accessible for use within the business.	Consumer information gathered from legacy systems and consumer-centric social applications, which are then integrated to create a complete view of consumers and the networks used by consumers. Consumers are seen as active participants in information updates through interaction-enabling technology.

Mangold and Faulds (2009:360) claim that consumer discussions have great power and are of critical importance to marketing strategies. Social media was perceived as an entertainment tool and transformed into a marketing instrument. Social media is now adopted as an essential part of a company's marketing combination. According to Cvijikj and Michahelles (2011:2) previous studies that have been done focus on the identification of an influential target group and the relationship between the identified target group and social media. These studies also address numerous challenges that are faced by social marketing.

One-to-one marketing has occurred, seeing that the consumer dictates his/her specific needs and requirements (Moncrief *et al.,* 2015:47). Salespeople find that the consumer is supplying information to the supplier and not the other way around. Social media marketing transforms the consumer from being a passive player in the buying process into an active participant by creating and sharing information and needs (Moncrief, 2015:47; Trainor, 2013:319). The consumer of the social media age can easily find needed information by making use of the Internet and social media. This results in a lack of face-to-face discussions between the consumer and the salesperson.

Customer relationship management has two objectives: firstly, customer retention through satisfied customers, and secondly, using customer insight for customer development (Tsiptsis & Chorianopoulos, 2009:1). It should be kept in mind that each customer is unique

and has different needs and behaviours. Customer relationship management is a two-stage concept: firstly, the basics of building customer focus should be mastered, and secondly, there is a move beyond the basics and development of customer orientation is integrated with customer relationship management (Rygielski *et al.,* 2002:492).

In the first stage (building of customer focus), the company should move from a product-orientation to a customer-orientation and should consider defining the market strategy from a customer's viewpoint (Rygielski *et al.,* 2002:492). In other words, the customer's needs are considered as more important than the features of the product. In the second stage (developing of customer orientation and integration thereof with customer relationship management), technology is used to achieve real-time customer management. In this stage, the value proposition of the company to the customers is constantly innovated (Rygielski *et al.,* 2002:492).

To summarize, customer relationship management is defined as a customer-centric strategy for building, managing and strengthening long-lasting customer relationships. Advantages of customer relationship management include detailed customer intelligence, which makes it possible to identify the most profitable customers, customer retention through satisfied customers and using customer insight for customer development.

Social customer relationship management is not a replacement for conventional customer relationship management. Social customer relationship management can be viewed as an extension of conventional customer relationship management. Advantages of social customer relationship management include transforming the consumer from being a passive player in the buying process into an active participant and making it easier to find needed information, not only for the customer, but also for the company.

This research aims to improve knowledge regarding techniques that can be used to identify the feelings, paradigms, sentiment and opinions of customers/clients. By identifying optimal techniques to do sentiment analysis and opinion mining, a company can gain more knowledge of the customer/client and can then transform a passive customer into an active customer. Subsequently, the components of customer relationship management will be discussed.

2.4.3 Components of customer relationship management

According to Rygielski *et al.* (2002:492), there are several components that are combined and form part of customer relationship management. Before data and text mining can be done, and customer intelligence can be gathered, the company must obtain customer information. Companies can either gather data internally or data can be purchased from sources outside the company (Rygielski *et al.*, 2002:492). Internal data sources can consist of (Rygielski *et al.*, 2002:492):

- Customer descriptions gathered from summary tables, for example a billing record of a customer.

- Some customers might find the time to respond to more detailed questions, such as a survey that is done.

- Web logs, credit card records, etc. gathered in transaction systems can provide behavioural data.

Today, most companies have one or more databases that contain marketing, human resources, and financial information. If a company wants to run a successful customer relationship management strategy, a critical component is an enterprise data warehouse (Rygielski *et al.,* 2002:492). The decision on whether to keep data granular or to aggregate data will depend on the level of experience in customer relationship management.

The customer relationship management system or strategy that the company is following should then analyse the data by using statistical tools, data-mining techniques and on-line analytical processing (OLAP) (Rygielski *et al.,* 2002:492). The company should consider employing data-mining analysts who are involved in the company and remind the company of the reasons why data-mining is being done (Rygielski *et al.,* 2002:492). If data-mining is done correctly, it will lead to successful market segments from which decisions can be made.

The last component that Rygielski *et al.* (2002:493) discuss is the campaign execution and tracking of the customer relationship management system. From data and text mining, decisions can be made and implemented through execution and tracking. Software to help marketing departments handle feedback from customers are still growing and being innovated (Rygielski *et al.,* 2002:493).

Campaign management software can also help manage and monitor customer communications across a variety of platforms. A company should also determine a

measurement process that can be followed to evaluate whether customer relationship management is done to the benefit of the company. In this study, data- and text mining techniques are identified, which can possibly increase the activity on a company's social media sites if the information gathered is transformed into new knowledge regarding customer/clients. Data- and text mining techniques used can also have an effect on the measurement process.

2.5   Social media measurement processes

Different techniques and strategies can be applied when a company decides to use social media, for example using a social media framework, choosing the most suitable social media platforms, ensuring that the content is timeless, using content in such a way that the company becomes the influencer, characterizing customers, using social media tools and templates, etc. After a comparison of techniques has been done, a company can then adjust its current social media strategy or framework and identify new opportunities and current weaknesses (AMEC, 2013).

A social media framework or model should reflect how social media is supposed to work with respect to the company and the consumers. A framework simply adds additional dimensions to the model, which is then measured with the use of metrics (AMEC, 2013). A good framework should also take into account and be flexible to numerous social media use cases. If a company sets measurable social media objectives, evaluation can be done to determine if the company is utilizing social media effectively. The type of company and industry will also have an influence on the measurement process that will be followed.

According to AMEC (2013), the following describes a general five-step social media measurement process or framework:
1. Firstly, the company should start with setting measurable objectives, which are aligned with desired business outcomes and Key Performance Indicators (KPIs).
2. The next step is to define specific metrics, which are necessary to measure and evaluate the performance against the measurable social media objectives that have been set.
3. The social media model should then be populated with the metrics that were selected within the framework.
4. After the measurement approach has been defined, data should be gathered and analyzed.

5. Lastly, the results should be reported to the stakeholders and the interested parties of the company.

The previously mentioned framework developed by AMEC (International Association for the Measurement and Evaluation of Communication) is still being researched and case studies have been done in which companies, such as Maybelline New York, Philips UK, Jaguar and Land Rover, etc. have been involved. This framework will be used as a guideline for this study during the creation of a social media framework, which can be used by companies (SO8).

Researchers have used the previously mentioned framework as a guideline and expanded the framework into an eight-step social media measurement process (Jeffrey, 2013; Paine, 2015):

1. Identify organizational and departmental goals of the company.
2. Research stakeholders for each of the set goals and prioritize the goals.
3. Set specific objectives for each prioritized stakeholder group.
4. Set social media KPI's (Key Performance Indicators) against each stakeholder objective.
5. Choose tools and benchmark (using a social media measurement matrix) public relations activity, intermediary effects and target audience effects.
6. Analyze the results and compare to costs.
7. Present to management.
8. Measure and evaluate on a regular basis to improve performance.

After a company has decided on the social media measurement process that will be used, the company should determine which social media platform's data will be used. The techniques that will be used to analyse the data also need to be determined. With the quantity of user-generated content being produced on social media platforms, it provides an excellent opportunity for data-mining, text mining and web mining (Dickinson & Hu, 2015:61). By using data-mining, text mining and web mining, a company can extract sentiment and opinions. This will be discussed in Chapter 3.

2.6   Conclusion

In this chapter, a literature review regarding the basic concepts that accompany the term social media, such as Web 2.0 and user-generated content was carried out. Web 2.0 tools are divided into five categories: blogs and podcasts, social networks, communities, content aggregators and virtual worlds. In this study, the focus will be on the use of social networks. Social networks offer the opportunity for a company to build a relationship with customers,

to follow and predict trends from the content generated by the user or customer, and to force innovation and re-evaluation of customer relationship management strategies.

Knowledge gained from this study will contribute to enabling a company to effectively utilize social media to the benefit of the company. In order to do this, user-generated content from customers that make use of external social media platforms needs to be extracted as data and then analyzed. Data will be analyzed to identify possible approaches for improving customer relationships. In this day and age, customers make use of more than one external social media platform, for example Facebook, Twitter, Google+, etc., therefore a company should consider the benefits of each external social media platform before spending resources and time on a social media platform that will not increase social capital.

To help manage and utilize customer relationships, social media data can be a powerful tool when analyzed. By analyzing and mining social media data, a customer's opinion or sentiment can be determined, as well as the influence that an expression can have on other users or customers. By using data- and text mining techniques, it can be applied to different aspects within the business'/company's environment. In Chapter 3, the emphasis is on how customer relationships can be improved by using different data- and text mining techniques to analyze social media data and to test the effectiveness of a company's social media sites.

CHAPTER 3: LITERATURE STUDY - MINING SOCIAL MEDIA DATA

3.1 Introduction

The aim in this chapter is to review how data- and text-mining techniques can be used to analyze social media data, with the aim of improving customer relationship management. Companies are constantly trying to strengthen their market position and gain a competitive advantage over companies with similar products or services in the industry. Companies are collecting social media data from user-generated content on a daily basis and analyzing this data can be beneficial to the company (Petre, 2013:21).

In section 3.2 of this chapter, different data-mining applications for businesses are discussed, as well as the role that data and text mining can play in improving customer relationship management. In section 3.3 the emphasis is on how social media data can be used by a company to conduct opinion mining and sentiment analysis. In this section, a distinction is also made between text mining, data-mining and web mining. In section 3.4, different techniques used in data and text mining are discussed and compared to identify the technique that are most suited for mining social media data.

Utilizing social media to the benefit of companies

Chapter 1: Problem statement

Chapter 2 - Literature Study: Social media	Chapter 3 - Literature study: Mining social media data
	3.1 Introduction
	3.2 Data and text mining in the business environment Data-mining applications for businesses are discussed, as well as how data-mining can be connected and used to improve customer relationship management.
	3.3 Using social media data for opinion mining and sentiment analysis Various terminologies associated with data-mining are discussed. Methods and techniques used in text mining are elaborated on.
	3.4 Data-mining techniques The following data-mining techniques are discussed: association rules (§3.4.1.), classification models (§3.4.2.), cluster analysis (§3.4.3.), multiple linear/logistic regression (§3.4.4.) and a comparison between data-mining techniques is also given (§3.4.5.).
	3.5 Conclusion

Chapter 4: Research design

Chapter 5: Results of questionnaire	Chapter 6: Results of Twitter data

Chapter 7: Discussion, interpretation and conclusion

Figure 3.1 - Overview of Chapter 3

In this chapter, a contribution is made towards finding possible answers to the research question, as the main objective of this study is to determine which techniques can be used by companies to effectively utilize social media and the data derived from social media platforms. The researcher aims to find an answer to the research question. The researcher reviews the identified techniques which can make it possible for a company to utilize social media more effectively.

In this chapter, a contribution is also made to the secondary objective (SO2): research different data- and text mining techniques that can be used to analyse social media data and gather social media data. How non-structured data (social media data), for example "tweets" of customers of a certain company can be captured, how patterns can be determined from the data, for example the identification of topic keywords and creating a connection between text sentiment and public opinion are also studied.

3.2   Data and text mining in the business environment

Data-mining is used to extract hidden information from not-structured or semi-structured data. According to Agrawal and Batra (2013:118), "data-mining is an important step of the knowledge discovery process". As mentioned in Chapter 1, a clear distinction can be drawn between data-mining, text mining and web mining. In text mining, the analyzer of the text applies certain aspects, techniques and methods of data-mining. Hotho *et al.* (2005:1) define text mining as "the process of extracting interesting information and knowledge from unstructured text."

Computers handle text as sequences of character strings (Hotho *et al.*, 2005:1). Text mining processes information and extracts meaningful numeric directories before the data can be processed through data-mining techniques, such as statistical and machine-learning algorithms (Agrawal & Batra, 2013:118). It must be kept in mind that user-generated content produced on social networks differs from conventional attribute-value data, which is used for classic data-mining (Gundecha & Liu, 2012:4). In section 3.2.1, different business areas in which data-mining techniques and methods have been applied successfully are discussed.

3.2.1   Data-mining applications for business

Petre (2013:23) describes three important business areas in which data-mining can successfully be applied: retail, banking and insurance. Figure 3.2 illustrates these three business areas and summarizes the applications of each business area. According to a broad classification model published by Statistics South Africa (2012:26), there are 21

sections whereby business areas in South Africa can be categorized. The broad classification model of sections/industries is given in Table 3.1.

Table 3.1 - Broad classification model of sections/industries (Statistics South Africa, 2012:26)

Section	Description/Category
A	Agriculture, forestry and fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity, gas, steam and air conditioning supply
E	Water supply; sewerage, waste management and remediation activities
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	Transportation and storage
I	Accommodation and food service activities.
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional, scientific and technical activities
N	Administrative and support service activities
O	Public administration and defense; compulsory social security
P	Education
Q	Human health and social work activities
R	Arts, entertainment and recreation
S	Other service activities
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
U	Activities of extraterritorial organizations and bodies, not economically active people, unemployed people etc.

From Table 3.1 the business areas which carry importance for this study can be categorized into two sections. Retail can be categorized into section G, while banking and insurance can be categorized into section K. In Figure 3.2 the chosen business areas are divided into segmentations in which data-mining can be applied successfully.
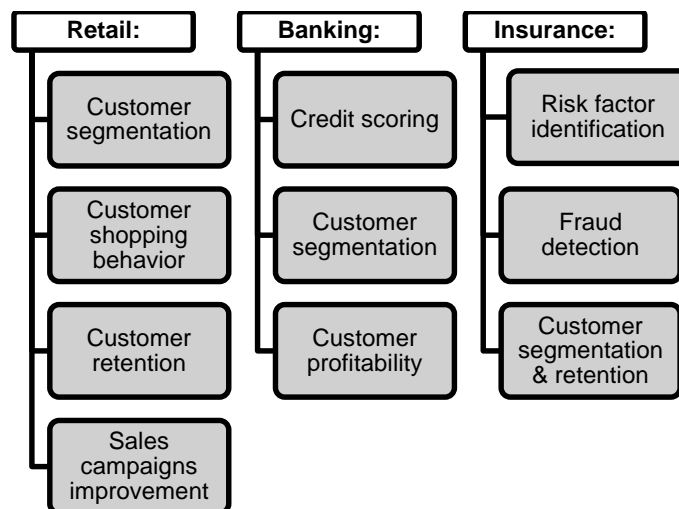
Figure 3.2 - Business areas in which data-mining can be applied successfully (Petre, 2013:23)

The data- and text mining techniques mentioned are a compilation from the mentioned sources and can be used for the following applications in the real industry (Petre, 2013:23; Rygielski *et al.,* 2002:488):

- *Customer segmentation:* In customer segmentation, different customer groups are identified and a customer is then associated to a corresponding group.
- *Discovering customer shopping behaviour:* Different customer buying patterns are identified and a forecast can be made regarding what product a customer will most likely buy next.
- *Customer retention:* In customer retention, the buying or shopping patterns are also identified and the product portfolio, prices and promotions can be adjusted.
- *Analyzing sales campaigns:* These applications allow a company to determine and predict how effective a sales campaign was or is going to be.
- *Performing basket analysis/affinity analysis:* This analysis identifies which products a customer is mostly to purchase together.
- *Sales forecasting:* Time-based patterns are identified, which assists retailers when making stock decisions.
- *Database marketing:* A company or retailer can determine which clients are, for example purchasing which labels or which clients attend sales. Database marketing also allows a company/retailer to plan cost-effective promotions.
- *Merchandise planning and allocation:* Retailers can identify which merchandise and what quantity should be placed in which demographic area when a new store is built in an area. Data-mining can also assist retailers in deciding on the layout of a store.

Operational and historical data is captured by information systems used by the banking industry. This allows data collected in this industry to be ideal for data-mining (Petre, 2013:24). The following is a combined list assembled from the mentioned sources and includes examples of applications in the banking industry (Petre, 2013:24; Rygielski *et al.,* 2002:489):

- *Credit scoring:* Factors are identified that can have an influence when a customer applies for a loan, credit card, etc.
- *Customer segmentation:* Similar to customer segmentation done in retail, clients are identified and a customer is assigned to a corresponding group. According to the assigned group, different banking options are offered to the client.
- *Customer retention/card marketing:* The product/service portfolio is adjusted according to the different transactions or banking trends that a customer follows.

- *Predict customer profitability/cardholder pricing and profitability:* Different patterns are identified, which are based on a variety of factors. This allows the profitability of the client/customer to be predicted.
- *Fraud detection:* A bank can identify fraudulent patterns by analyzing previous transactions of a possible client.
- *Predictive life-cycle management:* A client's lifetime value and value to service can be predicted.

Similar to the banking industry, the insurance industry also gathers large quantities of data that can be analyzed by using data-mining techniques (Petre, 2013:24). Data-mining techniques have the following applications in the insurance industry (Petre, 2013:24):

- *Risk factor identification:* This allows a company to identify factors that can have an influence on the client's level of risk and what the company is willing to insure according to the risk level.
- *Fraud detection:* Patterns of fraud can be identified, as well as the factors that may indicate a high probability of fraud for a client claim.
- *Customer segmentation and retention:* Customer groups are identified and a new client is added to the appropriate group. This allows a company to identify a package according to the group in which the client is placed and to increase customer loyalty.

Another application area where data-mining techniques are also used includes the telecommunications industry that analyzes call-detail records and predicts customer loyalty to the specific company (Rygielski *et al.,* 2002:489). Other applications assist in determining the number of customers who will submit a warranty claim. Airlines also make use of data-mining techniques to determine which customers qualify for incentives. This study investigates the value that data- and text mining of social media data can add to a company's customer relationship management and this is discussed in section 3.2.2.

3.2.2 Data-mining and customer relationship management

Customer relationship management is a broad topic that consists of many different layers. One of the layers discussed here, is data-mining (Rygielski *et al.,* 2002:493). If a company wants to evaluate the quality of its customer relationships, it can be done through data analysis. This can also contribute to understanding the needs of the customer better (Tsiptsis & Chorianopoulos, 2009:2).

Data-mining can also contribute to supporting customer development by introducing products to customers and contributing to product promotion campaigns. The customer lifecycle refers to the different stages and the relationship between the customer and the company. The customer lifecycle can be directly related to customer revenue and customer profitability (Rygielski *et al.*, 2002:493). Figure 3.3 is a representation by Tsiptsis and Chorianopoulos (2009:2) of how data-mining and the customer life-cycle management are integrated. The customer life cycle is important because it is a framework that allows a company to understand its customers' behaviour (Rygielski *et al.,* 2002:493).
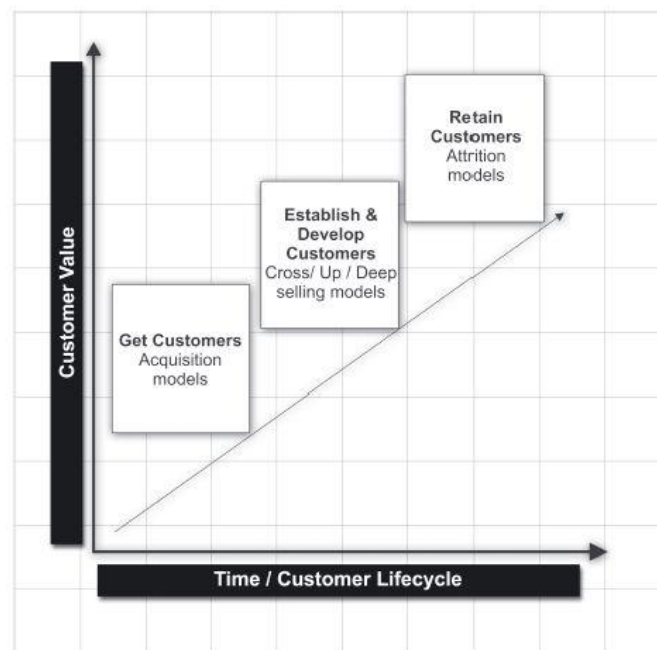
Figure 3.3 - The integration between data-mining and customer lifecycle management (Rygielski et al., 2002:493)

The customer life cycle consists of four key stages (Rygielski *et al.,* 2002:493):

1. *Prospects:* Prospects refer to prospective customers that are not yet customers, but form part of the target market.

2. *Responders:* Responders refer to prospects/prospective customers who have responded to the brand, product or service.

3. *Active customers:* Active customers are customers who are already using the product or service.

4. *Former customers:* Former customers include customers who no longer form part of the target market, are no longer using the product and started using a competing product or service, and customers who have not paid bills, and have incurred high costs.

The customers' use or purchases of the brand and products will increase if they have already used the product and found it impressive. Higher-margin products should be sold to

these customers and the customers' curiosity and interest in the brand or product should be kept for a longer period of time (Rygielski *et al.,* 2002:493). As the company evolves, the customer relationship will also change, because the company and customer will learn more about each other as times change.

When data-mining is used, reference is often made to the customer life cycle to help enhance customer relationship management (Rygielski *et al.,* 2002:493). The process of data-mining requires input and then delivers output. The input side of data-mining requires information and the customer lifecycle depicts what information is available. The output side of data-mining will allow the company to make decisions regarding product and customer information that have been derived from the data. The endpoint of the customer life cycle can inform the company what was found interesting during the time period of the cycle (Rygielski *et al.,* 2002:493). To improve customer relationship management, a company should be able to determine what their customers think, feel and how they react. By using data- and text mining techniques, the opinions and sentiment of customers can be determined.

To summarize, data-mining can be applied in different business areas. For this study, the banking, retail and insurance industries were identified as data-mining has already been successfully applied in these business areas/industries and these industries are focusing on the opinions of their customers/clients to help make better business decisions. The information retrieved from data-mining can contribute in improving customer relationship management. The research being done in this study will contribute to establishing and developing customers/clients as can be seen in Figure 3.3 (Rygielski *et al.*, 2002:493). In the next section, various terminologies that are associated with data-mining are discussed, as well as methods and techniques used to do data-mining and text mining. For this study, it is important to identify text mining techniques as the Twitter data that is used consists of text. Data-mining techniques can then be used to create valid assumptions from this text.

3.3   Using social media data for opinion mining and sentiment analysis

Pang and Lee (2008:1) define opinion mining and sentiment analysis as the computational treatment of opinion, sentiment, and subjectivity in text. Analyzing social media data for opinion mining and sentiment analysis is particularly important for this study because opinion mining and sentiment analysis will contribute towards improving the utilization of social media for a company. The information gained from the opinion mining and sentiment

analysis can be used as a generalization to improve customer relationship management discussed in Chapter 2 of this dissertation.

The main aim of opinion mining and sentiment analysis is to automatically extract opinions expressed in user-generated content created on social networks (Gundecha & Liu, 2012:5). By using data-mining techniques, opinion mining and sentiment analysis can be done. In Table 3.2, different terminologies associated with opinion mining and sentiment analysis are described. The different terminologies influence the data- or text mining technique that will be used.

Opinion mining and sentiment analysis tools are used by companies to understand product and service sentiments, as well as perceptions of users regarding a brand or a new product. Opinion mining and sentiment analysis tools also contribute to helping a company manage its reputation (Gundecha & Liu, 2012:5). Social media data is ideal for opinion mining and sentiment analysis because data gathered from social media networks can answer questions, such as how the user can be heard, which source of information the user should use, as well as how the user or consumer experience can be improved (Gundecha & Liu, 2012:4).

Table 3.2 - Terminologies associated with opinion mining and sentiment analysis (Adedoyin-Olowe et al., 2013)

Term	Definition
Sentiment lexicon	Dictionary containing sentimental words used by reviewers or clients on the organization's social media platform. Sentiment lexicon is a list of the common words that will enhance data-mining techniques when used in mining sentiment within a document.
Opinion definition and opinion summarization	An opinion definition can be found in a text, sentence or topic in a document. Opinion definition can also be allocated throughout the whole document. The sentiment polarities, degree and associated occurrence are analysed in opinion summarization. Opinion extraction is of critical value for opinion summarization.
Sentiment orientation	Sentiment orientation determines and labels the mood with no regard to the number of words or the opinion that is issued by the user of the social media platform.
Opinion extraction	The greater the number of people giving their opinion in a particular subject, the more valuable and important the portion of data will be for extracting.

In Figure 3.4, the major steps that are involved in sentiment analysis are illustrated. The data miner should firstly find the relevant documents and identify the sections that are relevant to the research question. This can be done by using text mining techniques (§3.3.2). When these two steps have been completed, the overall sentiment can be determined and quantified.
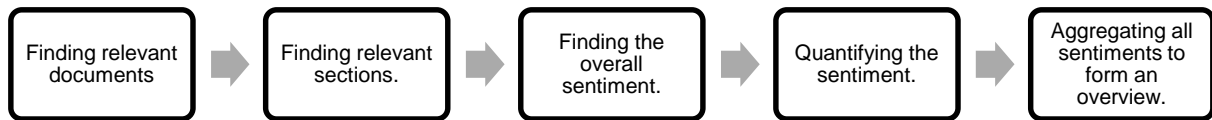
Figure 3.4 - The major steps of sentiment analysis (Gundecha & Liu, 2012:5)

Opinion mining tasks can be performed at document level, sentence level or feature level (during the first two major steps of sentiment analysis, Figure 3.3) (Gundecha & Liu, 2012:5). A difficult task for the data miner is to extract opinions, which are expressed in comparative sentences. The basic components of an opinion are illustrated in Figure 3.5.
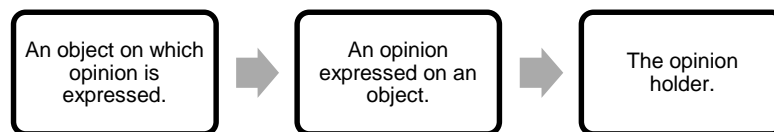
Figure 3.5 - Basic components of an opinion (Gundecha & Liu, 2012:5).

To conduct opinion mining and sentiment analysis text mining, data-mining or web mining is done on the unstructured data. To contribute to understanding text mining, data-mining and web mining the various associated terminologies, as well as methods should be discussed.

3.3.1 Various terminologies associated with data-mining

In this section, various terms related to data-mining, terms, such as text mining, Knowledge Discovery in Databases (KDD), Information Extraction (IE), etc. are described. The different methods used in text mining are briefly described. It is clear that there are different views on which methods can be categorized as text mining methods. In section 3.4, the techniques associated with data-mining are described. Data-mining techniques, which can be used to extract sentiment and opinion from social media data, will be identified.

Hotho *et al.* (2005:4) are of the opinion that the term text mining should be defined in the context with the related research area to which it corresponds. Text mining can relate to Information Extraction (IE), which is the extraction of facts from text (described in §3.3.2). Another field is data-mining, which Hotho *et al.* (2005:4) define as machine learning and statistics, which is used to find a beneficial pattern. There is a difference between statistical analysis and data-mining. Data-mining makes use of traditional statistical elements and pattern recognition (Moss & Atre, 2003:303). Table 3.3 indicates the differences between statistical analysis and data-mining.

Table 3.3 - Difference between statistical analysis and data-mining (Moss & Atre, 2003:304)

Statistical Analysis	Data-mining
1) A hypothesis starts the process.	1) No hypotheses are needed.
2) Equations are developed to match the hypothesis.	2) Data-mining algorithms can automatically develop equations.
3) Statistical analysis uses only numerical data.	3) Data-mining makes use of different types of data.
4) Statisticians can find and filter unstructured data during their analysis.	4) Data-mining makes use of structured data.
5) Statisticians interpret their own results and convey these results to the business managers and business executives.	5) Data-mining results are not easy to interpret. A statistician must still be involved in analyzing the data-mining results and conveying the findings to the business managers and business executives.

The last field, to which Hotho *et al.* (2005:4) relate text mining, is Knowledge Discovery in Databases (KDD), which is the discovery of information in large collections of data. Data-mining is thus regarded as an analysis step of the KDD process (Al-Azmi, 2013:5). The steps of KDD can be seen in Figure 3.6.

Figure 3.6 - Steps of the Knowledge Discovery in Databases process (Al-Azmi, 2013:5)

The term Business Intelligence (BI) has a direct relationship with data-mining. Golfarelli *et al.* (2004:1) define Business Intelligence as a process in which data is transformed into information and from the information knowledge is obtained. Knowledge regarding customer needs, customer decision-making processes, and competition in the industry, conditions of the industry, as well as trends that are being followed in the economy and technology can be obtained (Golfarelli *et al.,* 2004:1). Moss and Atre (2003:4) state that Business Intelligence is not a product or a system; it is a framework and a collection of integrated operational applications and decision-support applications that provide business data.

According to the two definitions, the term Business Intelligence (BI) can be defined as follows: Business Intelligence (BI) uses integrated operational and decision-support applications that consist of data-mining techniques to obtain information and to divert this information into knowledge that is beneficial to the business. The Knowledge Discovery Process (mentioned as KDD) is integrated in BI. In Table 3.3 a summary of the Business Intelligence advantages is given when data-mining methods and techniques are used by a company. The terms text mining and web mining also have a connection with data-mining.

Table 3.4 - Summary of Business Intelligence advantages when data-mining methods are used (Al-Azmi, 2013:14).

Business aspects	Business Intelligence (BI) advantages
Competitive advantage	- Market research→ elements of market dominance can be identified. - Risk management → bankruptcy prediction and better investment planning. - Optimizing manufacturing → Material usage is improved, as well as shipment and scheduling.
Customer Relationship Management (CRM)	- Targeting the customers that are most fitting to the target market. - Basket and market analysis → better marketing and advertising can be done. - Customer satisfaction → identifying the reasons and the costs accompanied with the satisfactory levels of customers. - Dynamic pricing.
Supply chain and logistics management	- Production management → Preventing over- and underproduction of products. - Can assist with reacting immediately to changes to help sustain supply. - Forecasting → Forecast the demand for production. - Help dynamically manage the supplies during their move through the chain.
Anomalies and fraud detection	- Fraud detection → Fraudulent transactions can be found. - Anomaly detection → Identify which data should be left out and why anomalies happened.

Kroeze *et al.* (2004:2) define text mining as a discovery of texts or the exploration of texts which contribute to the search of valuable, but hidden information. According to Al-Azmi (2013:2) text mining is an interdisciplinary field that encompasses "computational linguistics, statistics, and machine learning." Al-Azmi (2013:1) defines data-mining as the process of analyzing large databases, such as data warehouses or data found on the Internet to discover new information, as well as patterns and behaviours in the data. By analyzing multiple dimensions of datasets extracted from rational databases, a summary can be produced of general trends that are found in the dataset (Al-Azmi, 2013:2).

Another mining technique that is becoming more and more popular is web mining. Al-Azmi (2013:2) defines web mining as the automatic extraction of relevant information from the artefacts, activities, and hidden patterns that can be found on the World Wide Web (WWW). Web mining is usually used for tracking customers' online behaviour, for example tracking cookies, hyperlinks and traffic patterns of a web site (Al-Azmi, 2013:2).

In section 3.3.1.2 the different methods and techniques used in text mining are discussed. From the methods used for text mining, data-mining techniques were derived. The methods and techniques used in text mining serve as an introduction to data-mining techniques discussed in section 3.4. To be able to do sentiment analysis and opinion mining, text-mining techniques should first be applied to the identified text and then data-mining techniques can be applied to extract more information from the text.

3.3.2 Methods and techniques used in text mining

There is a variety of related research areas to text mining and text mining is defined according to each of these areas. According to Hotho *et al.* (2005:5), the following are text-mining methods and techniques and provide a structure in which data-mining methods can be categorized:

- *Information Retrieval (IR):* Information retrieval includes information processing, data retrieval and knowledge retrieval. Information retrieval is based on the idea of finding answers to questions that have been specified by a person or company.

- *Natural Language Processing (NLP):* Natural Language Processing focuses on understanding natural language and doing so by using computers and computational models. Techniques used in natural language processing include the manipulation of strings, automatic processing of natural language inquiries and linguistic analysis techniques.

- *Information Extraction (IE):* Information extraction is the extraction of specified information from text documents, which are then stored in patterns similar to the patterns of a database. Fan *et al.* (2006:78) define information extraction as a technology that represents a starting point for computers, which will help analyze unstructured text and with identifying key phrases and relationships within text.

Agrawal and Batra (2013:119) suggest that the previously mentioned related research areas can be separated into methods, which are used for either the mining of plain text or the mining of structured text. Table 3.4 indicates the methods or techniques used for mining plain text and mining structured text. The compiled list includes methods or techniques used when the input of text is in plain natural language (Agrawal & Batra, 2013:119; Fan *et al.,* 2006:79; Hotho *et al.,* 2005:5):

- Text summarization: A text summarizer will produce a compressed representation of the input that has been given and contains individual documents or groups of documents. Text summarization can help users figure out whether a rather long document is worth reading and will meet their needs. Related to text summarization is text compression. The output of text compression differs from text summarization as the output of text summarization is humanly legible whereas the output of text compression is not.

- Document retrieval: Document retrieval is frequently also referred to as information retrieval. Information retrieval is an extension of document retrieval. Document retrieval is the task of identifying and returning relevant documents.

- Information retrieval: As mentioned, information retrieval is an extension of document retrieval, where the documents are processed to extract specific information according to a query.

- Assessing document similarity: A problem in text and data-mining is the assessment of the similarity between different documents. An example is when documents are assigned to pre-defined categories and the grouping of documents into natural clusters can lead to a problem if documents are similar and it can then be problematic to extract the correct data and use data- and text-mining tools to their full potential.

- Text categorization: Text categorization consists of assigning pre-defined categories to natural language documents according to the content of the documents. The dominant approach when text categorization is being done is to make use of machine-learning techniques to gather categories automatically by using a training set of pre-classified documents. These categories are symbolic labels and have no other additional semantics. No other information, except the content of the document, is used when the document is placed into categories. The objective is to rank categories according to the relevance to a specific document. The labelling of categories can either be probabilistic or deterministic. When particular documents are categorized, a computer program often creates a "bag of words".

Table 3.5 - Methods/techniques used for mining plain text and structured text (Agrawal & Batra, 2013:119; Hotho et al., 1995:5)

	Text mining plain text	Text mining structured text
Methods/techniques used	- Text summarization, - document retrieval, - information retrieval, - information extraction, - assessing document similarity, - text categorization, and - Natural Language Processing (NLP).	- Wrapper induction, - document clustering with links, and - determining "authority" of Web documents.

Text from the Internet can contain explicit structural mark-up and is different from plain text (Agrawal & Batra, 2013:120). The term "Web mining" is becoming more and more popular and makes use of all the additional information which is available in Web documents. Agrawal and Batra (2013:120) and Kushmerick *et al.* (1997) label the following as text-mining methods, which can be used for structured Web text:

- Wrapper induction: A wrapper can be described as a function of a page. The page contains a set of tuples. Wrapper induction is a technique used to automatically construct a wrapper. This is done by performing queries on the set of tuples.

- Document clustering with links: Techniques used for document clustering are based on the similarity of text within a document.

- Determining "authority" of Web documents: A numeric rank for each web site is calculated. A web site might add a link to another web site, which is highly ranked, but the link might lead to a non-existent web site. Techniques are used by search engines to sort the rank of a web site according to a query by a user.

It should be kept in mind that text- and web mining form part of data-mining, even though there is a distinction between these three. Data-mining, text mining and web mining should not be seen as a single method or technique. Data-mining includes different approaches that search for patterns and relationships within data (Al-Janabi, 2013). Data- and text-mining techniques are of importance for this study as opinion mining and sentiment analysis (key-word based approach of text mining) will contribute to this study, but more detail regarding social media data is also required to improve customer relationship management and the utilization of social media by a company.

To summarize, opinion mining and sentiment analysis are used by companies to understand product, brand and service sentiments and opinions of customers/clients. There is a variety of terminologies that can be associated with data-mining and that can even be influenced when data-mining is applied, for example Business Intelligence (BI). Other terms that are also associated with data-mining include text mining and web mining. Data-mining techniques can also be applied after text mining and web mining have been done. One should not be of the opinion that statistical analysis and data-mining are the same. Data-mining uses statistical methods to make interpretations and valid assumptions of the data.

For this study, both text-mining and data-mining techniques will be used, these techniques will be used for the analysis of customers sentiment and opinions expressed on Twitter. Data-mining is ideal because the researcher does not have to have a hypothesis and different types of data can be used. In §3.3.2, different methods and techniques used in text mining are discussed. In the next section, different data-mining techniques will be discussed. In Chapter 5, the process, text-mining techniques, data-mining techniques and results gathered from the data used for this study will be explained in more detail.

3.4.   Data-mining techniques

Data-mining models and tools are created according to a set of rules, equations or complex transfer functions, which are then used to identify, understand and possibly predict data patterns from the data (Tsiptsis and Chorianopoulos, 2009:3). Social media data is vast, noisy, distributed, unstructured and dynamic, which may include challenges when the data is

mined (Gundecha & Liu, 2012:4). Before data can be effectively mined, the data should be structured and irrelevant data should be removed.

Social media data that is distributed has no central authority that maintains the data and because of this, new data-mining techniques and algorithms to be developed. In data-mining, it is necessary to pre-process the text. After data is extracted through information extraction methods, natural language processing or other pre-processing steps should be used and data-mining techniques should be applied (Hotho *et al.,* 2005:4). Figure 3.6 illustrates the main features of a data-mining solution for a company.
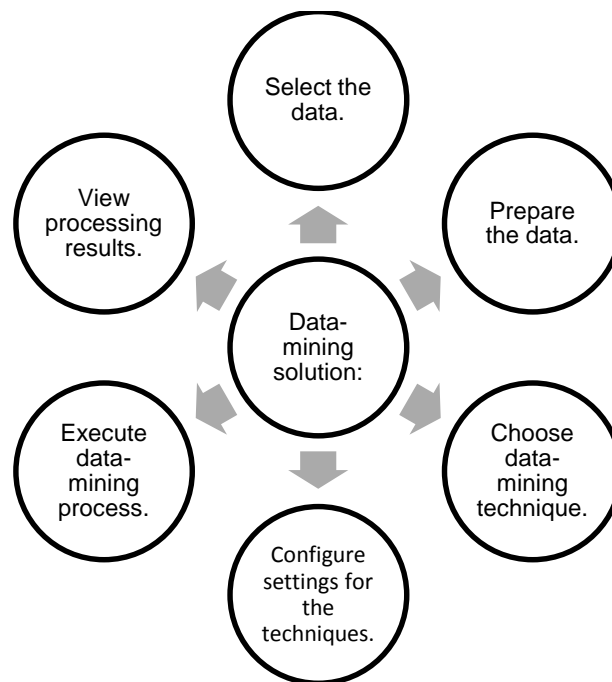
Figure 3.7 - Main features of a data-mining solution for a company (Petre, 2013:27)

Figure 3.8 illustrate the steps of data-mining.
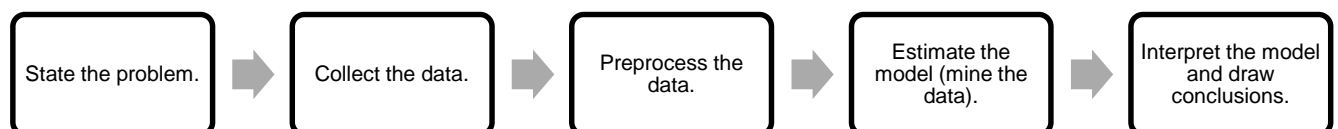
Figure 3.8 - Steps of data-mining (Hotho et al., 2005:4)

According to Tsiptsis and Chorianopoulos (2009:3), data-mining models can be grouped into two main classes according to their goal, namely supervised/predictive learning models and unsupervised learning models. The goal of supervised/predictive learning models is to predict an event or estimate the values of a continuous numeric attribute. These models contain an input field, as well as an output field. The input field is also referred to as a predictor, because this field is used by the model to identify a prediction function for the

output field (Tsiptsis and Chorianopoulos, 2009:3). Predictive/supervised learning models are further categorized into the following two models and described briefly (Tsiptsis and Chorianopoulos (2009:3):

1. Classification or propensity models: These models identify the target groups/classes from the beginning and the goal is to classify the cases into predefined groups, in other words, the model tries to predict an event.

2. Estimation models: Even though these models are similar to classification models there is one difference. These models are used to predict the value of a continuous field purely based on the observed values of the input attributes.

The other class, into which data-mining models are grouped, is unsupervised learning models. In these models there are no output fields; only input fields are used in these models and the pattern recognition is undirected. In other words, data patterns need to be uncovered in the set of input fields (Tsiptsis and Chorianopoulos, 2009:4). Unsupervised learning models can then be categorized into two models:

1. Association and sequence models: Association models identify associations between discrete events, products or attributes, whereas sequence models detect associations over a time period.

2. Cluster models: In these models the groups are not identified from the beginning and the data-mining algorithm will be used to analyse the input data, discover patterns and then identify natural groupings. When new groupings have been identified, they are scored and assigned to one of the revealed clusters.

Data-mining techniques and methods make use of Artificial Intelligence (AI) and neural networks (Al-Azmi, 2013:5). In sections 3.4.1 to 3.4.4, four main data-mining techniques and methods are discussed, as well as the algorithms that form part of a main data-mining technique and method. The four data-mining techniques being discussed are: association rules, classification models, clustering/cluster analysis and linear/logistic regression.

3.4.1 Association rules

Association rules, also known as relation rules, are the most commonly used data-mining technique (Brown, 2012:3). A correlation is made between two or more items, which identify patterns. Association rules are mostly used to do market-basket analysis (Hegland, 2001:317). A market-basket analysis is when a customer's basket of products or one transaction is analysed to see what the associated products that a customer will buy are. A basket/trolley will only cover a small portion of the products available, but it creates the

opportunity to observe which items are purchased in groups and to detect these groups (Hegland, 2001:317). This information is used to place products in specific demographical areas, and to improve marketing strategy. The principle and model of association rules are discussed next.

3.4.1.1   The principle and model of association rules

The data is modelled as a sequence of data records. The data records are the transactions that took place (Hegland, 2001:317). Agrawal *et al.* (1993:208) and Han *et al.* (2012:246) present the model as follows:

Let $I = \{i_1, i_2, ..., i_m\}$ be a set of binary attributes, called items. Let T then equal the transactions recorded in the database, and each transaction t is represented as a binary vector, with $t[k] = 1$ if t bought the item I_k, and $t[k] = 1$. Let X be a set of some items in I. The transaction t then satisfies X if for all items I_k in X, $t[k] = 1$. An association rule also can be described by the implication in the form of $X \Rightarrow I_j$. X is a set of some items in I.

I_j is a single item in I, this single item is not present in X. The rule $X \Rightarrow I_j$ is satisfied in the set of transactions T with a confidence factor $0 \leq c \leq 1$ if and only if at least $c\%$ of the transactions in T that satisfy X also satisfy I_j. The notation $X \Rightarrow I_j \mid c$ is used to specify that the rule $X \Rightarrow I_j$ has a confidence rule of c.

3.4.2   Classification models

Han *et al.* (2012:327) explain that classification can extract models that describe important data classes and these models are called classifiers. Classification can be used to create an idea, according to attributes, of the type of customer, item or object (Brown, 2012:4). Classification is a two-step process (Han *et al.,* 2012:327). In the first step, a classification model is built according to previous data that was gathered and the second step determines whether the model's accuracy is acceptable and whether the model can be used to classify new data (Han *et al.,* 2012:327). The first step is also known as the learning step and the second step as the classification step. Next, the principle and model of a classification algorithm will be discussed.

3.4.2.1   The principle and model of a classification algorithm

Han *et al.* (2012:328) present the classification algorithm as follows:

As mentioned, data classification is a two-step process: in the first step (learning step), the classification algorithm develops a classifier by learning from the training set that consists of database tuples and the tuples' associated class labels. A tuple is presented as X and is

represented by *n*-dimensional attribute vector, $X = (x_1, x_2 ..., x_n)$, *n* measurements made on the tuple form *n* database attributes $A_1, A_2 ..., A_n$.

An assumption is made that each tuple *X* belongs to a predefined class. The training set that consists of individual tuples, also known as training tuples, is randomly sampled from the database.

This step can also be depicted by the function *y = f(X)* that can predict the associated class label *y* of a given tuple *X*. Neelamegam and Ramaraj (2013:370) are of the opinion that the following algorithms are based on classification: decision trees (§3.4.2.1.1), k-Nearest neighbour (§3.4.2.1.2), support vector machines (§3.4.2.1.3), Naïve Bayesian classification (§3.4.2.1.4), and neural networks (§3.4.2.1.5). The listed algorithms are discussed in the subsequent sections.

3.4.2.1.1 Decision trees

Decision trees can be a powerful technique, not only for classification but also for prediction (Berry & Linoff, 2004:165). Decision trees represent rules, which make it simple to understand. A rule is represented as a path through the tree. Decision trees also make it possible to explore data while gaining insight into the relationships of large numbers of variables, varying from candidate input variables to a target variable (Berry & Linoff, 2004:165). Decision trees are used to divide a collection of records into smaller sets of records by applying decision rules into a simple sequence (Berson, 2005:22; Berry & Linoff, 2004:166).

There is a variety of decision tree algorithms, but the same basic process is followed. The data is continuously split into smaller groups, which ensures that each new generation of nodes is of greater purity than its parent nodes, while the target variable is kept in mind with each division (Berson, 2005:23; Berry & Linoff, 2004:171). The effectiveness of a decision tree can be measured by applying a test set of records. This test set is a collection of records that was not used to build the tree. The percentage of records that is then classified as correct is observed (Berson, 2005:23; Berry & Linoff, 2004:176).

This also provides a classification error rate of the tree and at each node the following can be measured (Berry & Linoff, 2004:176):

- the number of records entering a node, which can either be a leaf node or a branching node;
- the proportion of records in each class;

- how records can be classified if the records were to be depicted as a leaf node;
- the percentage of records classified correct at a specific node; and
- the variance in distribution between the set used to build the tree and the set used to test the tree.

Records are split at each node. This is done according to the function of a single input field. It should be decided which of the fields of input would deliver the best split (Berry & Linoff, 2004:172). The target variable can either be categorical or numeric (Berry & Linoff, 2004:178). Purity measures are used to evaluate splits for categorical target variables and the following measures are used: Gini index (population diversity), Entropy (information gain), information gain ratio and the Chi-square test (Berry & Linoff, 2004:178). There are two measures that can be used for numeric targets, namely reduction in variance and F test (Berry & Linoff, 2004:178).

Decision trees can be used in a variety of situations (Berry & Linoff, 2004:203). Decision trees can be used for data exploration to explore a large dataset and to determine which variables are useful. Decision tree methods can be applied to sequential events, which can help predict and simulate the future states of important variables.

3.4.2.1.2 k-Nearest neighbour

Nearest neighbour classifiers are based on learning by analogy; in other words, comparing a given test tuple with similar training tuples (Han *et al.,* 2012:423). Training tuples are described by an *n*-dimensional pattern space. The k-nearest neighbour classifier will search the pattern space if an unknown tuple is given to find the tuples that are similar or closest to the unknown tuple. The closeness is defined by a distance metric known as Euclidean distance and between two points or tuples ($X_1 = (x_{11}, x_{12} ..., x_{1n})$ and $X_2 = (x_{21}, x_{22} ..., x_{2n})$). This will be represented by the formula:

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2}$$

The k-nearest neighbour method is mostly used when predictions are made from text retrieval (Berson *et al.,* 2005:10). This method is found successful if the data is pre-formatted. When a prediction is made, some sense of confidence should be present to motivate a decision that is made from a prediction. If all the nearest neighbours make the same prediction, the confidence level is also considered to be higher (Berson *et al.,* 2005:11).

3.4.2.1.3 Support Vector Machines (SVM)

Han *et al.* (2012:408) describe a support vector machine as a classification method used for both linear and non-linear data. Wu *et al.* (2008:10) explain support vector machines as a classification function that distinguishes between members of two classes of the training data. A support vector machine makes use of non-linear mapping and transforms original training data into a higher dimension. In this higher or new dimension, a decision boundary that separates the tuples of one class from another is searched for., This separation is known as a hyperplane (Han *et al.,* 2012:408). The support vector machines method finds the hyperplane by making use of support vectors.

For a linear dataset, the linear classification function can be made to correspond to the hyperplane $f(x)$ that passes through the middle (Wu *et al.,* 2008:10). If the function is determined, a new data instance can be classified x_n. This can be done by testing if the sign of the functions $f(x_n)$ and x_n belong to the positive class if $f(x_n) > 0$. By maximizing the hyperplane between the two classes, the best function can be found. The reason why support vector machines need to find the maximum margin hyperplane is to deliver the best possible generalization ability. It also increases the classification performance of the method and ensures that classification can be done correctly on future data (Wu *et al.,* 2008:10).

To ensure that the maximum margin hyperplanes are found, the support vector machine classifier attempts to maximize the following function (Wu *et al.,* 2008:10):

$$L_P = \frac{1}{2} \; ||\vec{w}|| - \sum_{i=1}^{t} a_i y_i \left(\vec{w} \cdot \vec{x_i} + b \right) + \sum_{i=1}^{t} a_i$$

t is the number of training examples used and a_i where *i = 1,…,t* are non-negative numbers. L_P is known as the Lagrangian and a_i are the Lagrange multipliers. The hyperplane is defined by the vector $\vec{w}$ and the constant b.

3.4.2.1.4 Naïve Bayesian classification

According to Han *et al.* (2012:350), Bayesian classifiers are described as statistical classifiers that can predict class membership probabilities. This classification method is also based on Bayes' theorem and its performance is compared to decision tree and neural network classifiers. Bayes' theorem allows the calculation of the posterior probability $P(\text{H}|X)$ from $P(H)$, $P(\text{H}|X)$ and $P(X)$. Bayes theorem can be represented as:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Han *et al.* (2012:351) explain the model of this method as follows:

The principle and model of a classification algorithm described in §3.4.2.1 is also used in this model. There are *m* classes $C_1, C_2 ..., C_m$. Given a tuple *X,* the classifier will predict that *X* belongs to the class with the highest posterior probability, in other words, the naïve Bayesian classifier predicts that *X* belongs to the class C_i if and only if $P(C_i|X) > P(C_j|X)$ where $1 \leq j \leq m, j \neq i$.

The class C_i for which $P(C_i|X)$ is maximized and therefore known as the maximum posteriori hypothesis and the following equation is then derived from Bayes' theorem:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

3.4.2.1.5 Artificial neural networks

Neelamegam and Ramaraj (2013:373) state that an artificial neural network is often referred to as a neural network, because this computational model is based on the emulation of a biological neural network. Artificial neural networks predict models by using sophisticated pattern detection, as well as machine-learning algorithms (Berson *et al.,* 2005:28). Artificial neural networks can also create complex models, which are not always the easiest to fully understand. A neural network requires a number of parameters that are determined empirically (Han *et al.,* 2012:398).

Many different neural networks and neural network algorithms exist, and the most used neural network algorithm is backpropagation (Han *et al.*, 2012:398). The backpropagation neural network algorithm performs learning on a multilayer feed-forward neural network. An artificial neural network is adaptive and can make changes to the structure of the network according to the external and internal information that is placed through the network during the learning phase (Neelamegam & Ramaraj, 2013:373). This type of algorithm iteratively learns a set of weights for prediction of the class label or the tuples and consists of an input layer, one or more hidden layers and an output layer (Han *et al.*, 2012:398).

3.4.3 Cluster analysis

According to Han *et al.* (2012:444), cluster analysis, also known as clustering, is a process whereby a set of data objects is divided into subsets. Each subset is then known as cluster and the objects within the cluster are similar to one another. If different clustering methods are used different clusters may be generated from the same data set. Clustering is considered useful because it brings forth unknown groups from the data set (Han *et al.,*

2012:444). Clustering or segmentation is done to give an end user a high level view of what can be depicted from the data (Berson *et al.,* 2005:12).

Business intelligence, image pattern recognition, Web search, etc. make use of clustering methods. For example, a large number of customers can be organized into different groups based on similar characteristics, which can be beneficial for business intelligence (Han *et al.,* 2012:444). This can help develop business strategies, which can lead to enhancements in customer relationship management (Han *et al.,* 2012:44). The k-nearest neighbour method, described in §3.4.2.1.2, used in classification is often compared to clustering. The k-nearest neighbour algorithm is a refinement of clustering (Berson *et al.,* 2005:13).

A distinction is made between these two methods; clustering is an unsupervised learning technique and the k-nearest neighbour is a supervised learning technique (Berson *et al.,* 2005:15). A supervised learning technique will try to perform a prediction, whereas with an unsupervised learning technique there is no reason for the creation of the model. Hierarchical and non-hierarchical clustering are known as the main types of clustering techniques.

3.4.3.1   Hierarchical clustering

When hierarchical clustering techniques are used, a hierarchy of clusters from small to big is created (Berson *et al.,* 2005:17). The user can determine what the correct number of clusters is that summarizes the data and provides useful information. To build a hierarchy of clusters, there are two main types of hierarchical clustering algorithm: agglomerative and divisive (Berson *et al.,* 2005:18).

When using an agglomerative clustering technique, each record is started off as a cluster, and the clusters that are the closest are then merged together. This process continues until all the records are at the highest level in the hierarchy and a cluster has been formed (Berson *et al.,* 2005:18). Divisive clustering techniques are the opposite of agglomerative clustering techniques. All the records in one cluster are started with and the cluster is then split into smaller pieces. Agglomerative clustering techniques are used in most situations (Berson *et al.,* 2005:18).

3.4.1.2   Non-hierarchical clustering

Non-hierarchical clusters can be created faster than hierarchical clusters, but these techniques require that the number of clusters should be decided on by the user (Berson *et*

al., 2005:18). Non-hierarchical clustering is run multiple times and clustering is done randomly as the technique is started and is then iteratively improved and clusters are formed as data is shuffled. The two main non-hierarchical clustering techniques are single pass methods and reallocation methods (Berson *et al.,* 2005:19). Single pass methods pass once through the data and clusters are then formed. Reallocation methods move records and data from one cluster to another cluster, which is more suited.

3.4.5 Multiple linear/logistic regression

The most popular mathematical model used to make predictions is linear/logistic regression (Berson *et al.,* 2005:6). Linear/logistic regression can be used in various data-mining situations, for example predicting the activity on a credit card by using demographical information and historical activity patterns, the expenses that accompanies travel or the vacation of a customer can also be predicted based on historical data and sales information, the prediction of sales and the impact that discount can have on sales in retail outlets. When the linear/logistic regression model is used the data needs to be split into two categories; firstly, the training data set, and secondly, the validation data set, which will be used to validate the multiple linear regression model (Berson *et al.*, 2005:6).

3.4.5.1   The principle and model of linear/logistic regression

According to Ye (2003:168), a linear regression model can be formulated as:

$$f(\boldsymbol{x}) = \ \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_d x_d = \ \beta' \boldsymbol{x}$$

$d + 1$ is described as the dimension of the feature vector $\boldsymbol{x}$, and the first element of $\boldsymbol{x}$ is 1. This result in not starting with an unspecified $f(\boldsymbol{x})$ and the model $d + 1$ can be used to estimate parameters. The coefficient β_j is used to represent the difference between two subjects that have the same feature vector, but differ by 1 unit on feature x_j. Linear regression models are fitted with a categorical feature with k levels.

The model described is used for linear regression. When a 0 or 1 classification problem needs to be solved, linear logistic regression is used. According to Ye (2003:169), statistical methods used for binary classification are designed in a way to produce estimates of $p(\boldsymbol{x}) = P(Y = 1|\boldsymbol{x})$. A linear logistic regression model assumes a parametric form for $p(\boldsymbol{x})$. Similar to linear regression, the prediction depends on the feature vector and the linear combination of the components of $\boldsymbol{x}$.

$$p(\boldsymbol{x}) = \frac{1}{1 + \exp(-f(\boldsymbol{x}))}, \qquad where\ f(\boldsymbol{x}) = \ \beta' \boldsymbol{x}$$

If the function is rewritten, it can be seen that the model assumes that the log-odds are a linear function of x. The logistic function transforms the linear combination to the [0, 1] interval.

$$log \frac{p(x)}{1 - p(x)} = \beta' x$$

3.4.6 Comparison between data-mining techniques

The sentiments and opinions expressed by users of social media platforms can affect the decisions that other users or customers of an organization make (Adedoyin-Olowe *et al.*, 2013). Sentiment analysis and opinion mining is of importance to this study seeing that it can have an influence on building and managing customer relationships. In order to determine what influence sentiments and opinions of users or customers have, data-mining techniques need to be applied to social media data.

Data-mining techniques can handle three problems associated with social media data. These problems include size, noise and dynamism (Adedoyin-Olowe *et al.*, 2013). Data sets of social media platforms are vast and to process and analyse these data sets within a reasonable time requires automated information processing. Spam blogs and irrelevant tweets, in the case of Twitter, can be described as noisy data. Social media data is dynamic as it is rapidly changing (Adedoyin-Olowe *et al.*, 2013).

A comparison of data-mining methods and the algorithms associated with each of the methods is depicted in Table 3.5. Previous research that has been done indicates that unsupervised learning methods are suitable when mining social media data (Adedoyin-Olowe *et al.*, 2013; Nandi & Das, 2013:165). Supervised learning methods are also suitable, but it depends on the results that the researcher wants to generate from the social media data.

Table 3.6 - Comparison between data-mining methods

Data-mining method	*Algorithms associated with data-mining method*	*Supervised/ Unsupervised learning method*
Association rules	- Apriori algorithm, - equivalence class transformation algorithm, and - frequent pattern growth algorithm.	Supervised learning method
Classification models	- Decision trees, - k-nearest neighbour, - Support Vector Machines, - Naïve Bayesian classification, and - artificial neural networks.	Supervised learning method
Clustering	- Partitioning algorithm, - hierarchical clustering, - density-based algorithm, - grid-based algorithm,	Unsupervised learning method

Data-mining method	Algorithms associated with data-mining method	Supervised/ Unsupervised learning method
	- model-based algorithm, and - constraint-based algorithm.	
Linear/logistic regression	This method follows the model explained in §3.4.4.1 to make predictions, interpretations and advantages.	Supervised learning method

By using clustering methods, users that depict the same opinion are linked to the set of nodes, and users that express an opposing opinion are then linked to another set of nodes. Clustering techniques are used in sets of data where the basis of the data has already been established, but the data pattern is unknown (Adedoyin-Olowe *et al.*, 2013). Supervised learning methods, such as classification and association can also be used. These methods will be used when the data organization is already identified (Adedoyin-Olowe *et al.,* 2013).

The data-mining specialist using supervised learning methods needs to understand the problem that needs to be solved and which solutions can be applicable to the problem. The most used algorithms include association rule mining, decision trees, k-nearest neighbour and neural networks, but algorithms that are becoming more popular include support vector machines and the naïve Bayesian algorithm (Adedoyin-Olowe *et al.,* 2013). New data-mining techniques will be developed as the data on-line and on social media platforms are ever increasing. The researcher will make use of a supervised learning method. The data-mining method/technique that is most suited for this study is classification models.

3.5   Conclusion

This chapter describes different data-mining methods, including association rules, classification models, clustering techniques and linear/logistic regression. Association rules identify frequent patterns, association, correlations or structures within the data. Classification is the process of organizing and categorizing data into categories identified in the data. Clustering is also a machine-learning technique that places data into related groups without the data-mining expert having advance knowledge of the group definitions. Regression predicts the probability of an outcome that is determined from the data.

Data-mining techniques are capable of mining social media data to extract user opinions and sentiments (Adedoyin-Olowe *et al.*, 2013). It is thus the opinions and sentiments that can be extracted, which characterize social media data. Social media data can provide vital information for a company and this information can be used as a decision support mechanism.

Text mining discovers or creates new knowledge from a collection of documents. Text-mining terminologies and techniques are often incorporated into data-mining. In this chapter, a contribution is made to improving the awareness and knowledge regarding the importance and value of social media if analyzed correctly. In this way, one of the secondary objectives of the study (SO2) is achieved, as the literature discussed in this chapter will contribute towards identifying which data-mining techniques and methods can be used to determine patterns from a company's social media platforms and other user-generated content.

Sentiment analysis and opinion mining can contribute to understanding the customer, which can improve the customer relationship management strategies used by a company. For this study, sentiment analysis and opinion mining will be done on Twitter data. The messages posted on Twitter are very short and less than 140 characters. Twitter data was chosen because the process to collect the data is easy without additional costs. Day-to-day sentiment scores will be derived by dividing the messages, known as tweets, into three categories, namely positive, negative and neutral.

A tweet is defined as positive if any positive words are included within the tweet and negative if any negative words are included. By dividing the tweet into positive or negative, sentiment can be determined. The tweet can also be further divided into subjective or objective, which indicates the opinion of the customer. The data-mining method that will be used for extracting sentiment and opinions will be a classification model.

A classification model is used because a tweet can take on a categorical value; positive, negative or neutral. By using classification, an assumption is made that the tweet expresses sentiment and opinions of a single entity (customer). The algorithms used in a classification model are best suited for aggregating the sentiment and opinions of each tweet after sentiment analysis and opinion mining have been done. For this study, a decision tree, Support Vector Machine (SVM), and Naïve Bayesian classifier, will be used. A discussion follows in Chapter 4 regarding the gathering of the Twitter data and in Chapter 5 the use of the data-mining method will be discussed.

CHAPTER 4: RESEARCH DESIGN

4.1 Introduction

In this chapter, different research paradigms and methods are discussed and the research paradigm that is most suited to this study is identified. The research design supports the researcher in deciding on the logic and structure of the research and delivers evidence necessary for answering the research question as accurately and clearly as possible. The framework or research design followed throughout the research influences the quality of the research (Dalcher & Brodie, 2007).

In section 4.2, three different research paradigms are described, namely the positivist, interpretivist and critical research paradigms. This section elaborates on the research paradigm that is used in this study. The research design should not be confused with the choice of research method. In section 4.3, different research methods associated with the chosen research paradigm are described. The data generation method used for this study is described in section 4.4, and data analysis methods in section 4.5 of this chapter. Figure 4.1 depicts the order of this chapter.

Utilizing social media to the benefit of companies

Chapter 1: Problem statement

| **Chapter 2 - Literature Study: Social media** | **Chapter 3 - Literature study: Mining social media data** |

Chapter 4: Research design

4.1 Introduction

4.2 Research paradigms
The following research paradigms are discussed: positivist research paradigm (§4.2.1), interpretive research paradigm (§4.2.2) and the critical research paradigm (§4.2.3). The research paradigm chosen for this study is the positivist research paradigm and requires the researcher to acquire knowledge objectively by using scientific methods.

4.3 Research method

4.4 Data-generation method
Questionnaires and social media data from external platforms such as Twitter data (tweets) will be used as method of data generation.

4.5 Data analysis
Graphical data analysis methods (tables, charts, and graphs), computer-aided analysis techniques, statistical techniques and data-mining techniques will be used to analyze the gathered data.

4.6 Conclusion

| **Chapter 5: Results of questionnaire** | **Chapter 6: Results of Twitter data** |

Chapter 7: Discussion, interpretation and conclusion

Figure 4.1 - Overview of Chapter 4

4.2   Research paradigms

According to Oates (2006:13), a paradigm can be defined as a pattern or shared way of thinking. Smyth and Morris (2007:424) define a paradigm as an intellectual framework, which embodies a person's ideas and beliefs. In this chapter, a paradigm will be defined as a person's view of the world and the beliefs that a person develops regarding certain aspects. According to Guba and Lincoln (1994:107), paradigms are based on ontological, epistemological and methodological assumptions.

Guarino (1995:628) defines ontology as the study of the organization and the nature of the world that is independent from a person's knowledge about the world. According to Krauss (2005:758), ontology involves the philosophy of reality. Ontology can be described as the assumed nature of reality being studied as realist or idealist (Hallebone and Priest, 2009). Epistemology, in a sense, is related to ontology and methodology. Epistemology addresses how a person came to know of a reality while methodology identifies the practices that were followed to attain knowledge (Krauss, 2005:758).

According to Krauss (2005:758), the term epistemology is derived from the Greek word *epistêmê*, which is Greek for knowledge. Guarino (1995:628) defines epistemology as knowledge that consists of propositions, formal structure of which is the source of new knowledge. Krauss (2005:759) is of the opinion that epistemology poses questions that are used by the researcher to try and determine what knowledge is, how this knowledge has been obtained and what counts as knowledge.

Ontology and epistemology are both sub-branches of philosophy. Ontology can briefly be described as the nature of being, reality or phenomena (Hallebone & Priest, 2009:48). A question associated with the term ontology is: What constitutes reality and what is the nature of the world? Epistemology, on the other hand, can be described as the theory of knowledge (Hallebone & Priest, 2009:45). The following question can be associated with epistemology: What establishes valid knowledge and how can this knowledge be obtained?

Ontology lays the foundation for epistemology, as ontology creates a need to know about the nature of being, reality or phenomena. An interpretation that can be made is that epistemology then builds on ontology by studying or striving to know how the being, reality or phenomena exist and what they are. In this dissertation, three research paradigms are discussed, namely the positivist, interpretive and critical research paradigms with reference to the ontology and epistemology of each research paradigm.

4.2.1 Positivist research paradigm

A researcher who chooses to follow a positivistic approach assumes that the reality is objectively given (Myers, 1997). This reality can be described by measurable properties which are independent of the researcher. The only way that knowledge can be observed as true is if it was created by using scientific methods, and data is derived from experiment and observation (McGregor and Murnane, 2010). From these definitions, the positivistic research methodology can be defined as a paradigm, which is based on the scientific method.

According to Oates (2006:283), the scientific method has two basic assumptions; firstly, the world we live in is ordered and regular, not random, and secondly, the world we live in can be investigated in an objective manner.

Before describing the characteristics of the positivistic paradigm, the basic beliefs of this paradigm need to be made clear. The positivistic paradigm has the following beliefs:

- Ontology: Undeniable natural laws and mechanisms are the driving force of an apprehensible reality. Knowledge of how the world is, is summarized in the form of time-free and context-free generalizations, some of these generalizations take the form of cause-effect laws (Guba & Lincoln, 1994:109; Orlikowski & Baroudi, 1991:12). The "true" state of affairs can be converged by research. The basic posture of this paradigm can be argued to be both reductionist and deterministic (Guba & Lincoln, 1994:109; Hesse, 1980; Orlikowski & Baroudi, 1991:12).

- Epistemology: The positivistic research paradigm can briefly be described as dualistic and objective. The researcher and the "object" being researched are assumed to be independent entities of each other. This allows the researcher to investigate the object without having an influence on the object or being influenced by the object (Orlikowski & Baroudi, 1991:12). Threats to the validity of the research project are recognized when the researcher or the object have an influence on one another. According to Guba and Lincoln (1994:110), various strategies are followed to ensure that this does not happen or to reduce and eliminate these threats. If the findings are repeated, and the same results are found, the findings are claimed as "true".

Oates (2006:286) is of the opinion that people who follow a positivist paradigm, work in this paradigm because it has the following characteristics:

- The world that exists is independent of humans: The world that exists is a physical, as well as a social world; this world does not only exist in a person's mindset and should be studied, captured and measured.

- Measurement and modelling: By making observations, measurements and producing models, such as hypotheses and theories of how the world works, the researcher will discover this world. A correspondence between *the researcher's model and* the features of interest in the world will be revealed. Oates (2006:286) explains that there will be one model or explanation for any aspect of the world, which can then be interpreted as "the truth".

- Objectivity: The researcher will act neutrally and objectively as an impartial observer. The researcher's personal values and beliefs will not affect any discoveries made.

- Hypothesis testing: The research being done is based on empirical testing of theories and hypotheses, which leads the researcher to the confirmation or refutation of these theories and hypotheses.

- Quantitative data analysis: The use of mathematical modelling and proof is strongly preferred because mathematics provides a logical and objective means for analysing observations and results.

- Universal laws: Universal laws, patterns and irrefutable facts are looked for to show that they can be true, regardless of the researcher and the occasion, and to find if there are any generalizations.

A positivistic paradigm makes use of the scientific methodology, which will allow the subject being studied and the researcher not to influence one another and the subject is measured against universal laws. Positivistic researchers collect scientific data that is precise and can be interpreted as quantitative data, which can be analysed by making use of statistics and mathematical techniques. When the data is analysed, the researcher has the intention to find patterns and to generalise the findings according to these patterns. The aim that needs to be achieved by the research is to explain or measure the research statements against reality.

4.2.2 Interpretive research paradigm

According to Oates (2006:292), interpretive research can be defined as follows: interpretive research in information systems and computing is concerned with understanding the social context of an information system, the social processes by which it is developed and construed by people and through which it influences, and is influenced by, its social setting. Orlikowski and Baroudi (1991:5) are of the opinion that interpretive studies assume that people create and associate their own subjective and intersubjective meanings while interacting with the world around them. Researchers using the interpretive paradigm attempt

to understand a certain phenomenon by accessing participants involved in the study (Klein & Myers, 2001:220; Orlikowski & Baroudi, 1991:5).

The term interpretivistic research is defined as: a mutual recognition between the researcher and the research participants are fostered and valued (Weaver & Joanne, 2006:461). According to Mackenzie and Knipe (2006), the following words can be associated with the interpretivistic research: naturalistic, phenomenological, hermeneutic, ethnographic, multiple participant meanings, social and historical construction, theory generation and symbolic interaction.

The interpretive researcher should be able to convince an audience that the descriptions, explanations and interpretations are plausible and can be sustained by the data collected (Oates, 2006:295). The interpretivist researcher relies on the participants' views of the situation being studied, while recognising the impact of the researchers own background and experiences on the study (Mackenzie & Knipe, 2006). Interpretivistic researchers do not begin with a theory; a theory is generated or inductively developed (Mackenzie & Knipe, 2006).

Travis (1999) summarizes the interpretivistic research methodology according to the ontology, epistemology and methodology as follows:
- Ontology: In interpretivism, realities exist in the form of multiple mental constructs, socially and experientially based, local and specific, dependent on their form and content on the persons who hold these constructs. Interpretive researchers study meaningful social action and the collect large quantities of data, mainly in the form of qualitative data, to construct an in-depth understanding of how a participant/person creates a meaning regarding a specific aspect of the world.
- Epistemology: The interpretive researcher is subjective. The research findings are the creation of the interaction between a participant in the study and the researcher. The researcher is subjective and works from a realised bias, while considering certain ethical concerns.
- Methodology: The individual constructions are elicited and refined, while using a hermeneutically approach. This means that the individual constructions are compared and contrasted while the researcher aims to generate one or more social constructions. The researcher aims to also generate a shared understanding.

Researchers working within the interpretivistic paradigm tend to have the same world view and the following characteristics are identified by Oates (2006:292):

- Multiple subjective realities: This characteristic claims that there is no singe version of "the truth". What the researcher and participants of the study claim to be 'real' or 'knowledge' is what is constructed in their minds, either individually or in a group. A person's culture can also influence the way the world is perceived.

- Dynamic, socially constructed meaning: Reality or whatever is perceived as reality can only be accessed and transmitted to others by creating more social constructions, for instance language, shared meanings and understanding.

- Researcher reflexivity: This characteristic claims that the researcher making use of the interpretivistic research paradigm cannot be viewed as neutral. The researcher's own assumptions, beliefs, values and actions will have an effect on the research process and the situation. If a researcher decides to use this paradigm, the researcher should be reflexive or self-reflective. This means that the researcher should acknowledge his/her influence on the research and how his/her interactions with participants in the study can lead to the renegotiation of meanings, understanding and practices.

- Study of people in their natural social setting: The research aims to understand an object or a person in their own world. The natural setting is studied by the researcher as well as other participants, without imposing on the object's or person's previous understanding and expectations of a situation.

- Qualitative data analysis: The researcher prefers the generation and analyses of qualitative data.

- Multiple interpretations: After the study has been conducted, more than one explanation is discussed by the researcher; this is because more evidence can be generated when qualitative data is used. Researchers expect to arrive at more than one fixed explanation.

Joubish *et al.* (2011:2083) imply that qualitative research is a generic term for investigative methodologies. Qualitative research and data cannot always be illustrated as quantities, for instance a person's beliefs, meanings, attributes, symbols, and other social or human problems (Joubish *et al,* 2011:2083). This implies that the study might involve content which needs to be analysed. The characteristics of qualitative research should not be forgotten, and also have an influence on the characteristics of the interpretive research methodology.

Joubish *et al.* (2011:2083) identify the following characteristics of qualitative research:

- Purpose: The purpose of qualitative research is to understand a person's interpretation.

- Reality: People's perceptions of reality change and therefore qualitative research should be dynamic.

- Viewpoint: Reality is what a person perceives it to be, therefore the viewpoint of the insider is very important.

- Values: When conducting and reporting on qualitative research, the researcher should take into account that the values do have an impact and should be understood.

- Focus: Qualitative research has a holistic view, in other words a total or complete picture.

- Orientation: Qualitative research allows the researcher to discover and form theories and hypotheses as data is collected.

- Data: Data collected is mainly about the perceptions of people in a certain environment and the researcher should be subjective at all times.

- Instrumentation: The human person is the primary collection instrument.

- Conditions: Investigations are conducted by the researcher while the participant is under natural conditions.

- Results: The focus is on design and procedures to gain "real", "rich" and "deep" data that can be claimed as valid.

The characteristics of the methods used in interpretivistic research are well suited to the characteristics of the paradigm itself, which ensures that the researcher observes the setting/environment of the participant objectively, while keeping in mind that the researcher's own paradigms and experiences should not have an effect on the study. The aim of this research paradigm is to understand interpretations of the research and the world/reality in which the interpretations where made.

The researcher should also consider doing an audit trail of the research, which will ensure that the researcher's own perceptions and paradigms did not influence the study, and that interpretations made about the data are correct. Qualitative data, as well as quantitative data can be used in interpretivistic research. The correct data-collection techniques, as well as data-analysis techniques should be used to ensure that interpretations made can be supported by the data.

4.2.3 Critical research paradigm

According to Oates (2006:296), critical research in information systems is concerned with identifying power relations, conflicts and contradictions, and empowering people to eliminate them as sources of alienation and domination. Critical research can also be defined as

research characterized by an intention to change the status quo, to overcome injustice and alienation, and to promote emancipation (Stahl, 2008:139). Critical researchers are of the opinion that social reality possesses objective properties, which tend to dominate a person's experiences and the paradigm that a person forms of the world (Oates, 2006:296).

According to Brooke (2002:49), critical theory is described as a form of historical materialism, which is influenced by issues, such as class, ethnicity and gender. When critical research is used for the study of a situation/subject, it is kept in mind that the situation's, subject's or participant's mindset might be dominated by another subject or participant (Brooke, 2002:49). The researcher, considering the critical research methodology, should understand the basic beliefs and characteristics of this methodology to ensure that ethical issues are addressed correctly in a study (Brooke, 2002:50).

Guba and Lincoln (1994:109) summarize the basic beliefs of the critical theory as follows:

- Ontology: Historical realism refers to virtual reality shaped by social, political, cultural, economic, and ethnic and gender values, which are crystalized over time. For all practical purposes, the structures used in this research paradigm are considered "real", a virtual or historical reality.

- Epistemology: The critical theory has a transactional or subjective nature in which mediated findings are valued. The researcher and the object that is being researched are assumed to be interactively linked, with the values of the researcher inevitably influencing the inquiry. Findings are therefore value-mediated.

- Methodology: The transactional nature of inquiry requires a dialogue between the researcher and the participants of the research; this indicates that methods and processes used are dialogical and/or dialectic.

Sudersan (1998:256) distinguishes between traditional theory and critical theory. The following characteristics of critical theory are highlighted:

- Critical theory states that there is no absolute knowledge of the subject being studied and that the coincidence of the subject and object lies in the future. This is not only due to intellectual progress, but also social progress in which the relationship between the subject and the object being studied is redefined.

- The method of sciences is different, because the end determines the means which lead to the means and the ends of critical theory being equally important.

- A critical reflection on ideology is used and the researcher accepts that this method of research is historically grounded, and the methodology cannot set itself free from any influences in a social framework.

- Critical theory realizes the importance of praxis and reposes faith in ideals of freedom, justice and happiness.

Brooke (2002:50) connects the words emancipation, power and resistance to critical research. According to Oates (2006:297), there are different styles and types of critical research, and there are five common themes that are identified:

- Emancipation: Critical researchers are dedicated to releasing people from the power that shapes organizations and society in the world known today. Critical researchers try to empower people and do not just try and explain a person's behaviour or culture. This research paradigm is therefore more activist than paradigms such as the positivistic research paradigm and the interpretivistic research paradigm.

- Critique of tradition: Critical researchers tend to question and challenge the status quo. Critical researchers highlight and confront existing patterns of power and taken-for-granted assumptions.

- Non-performance intent: Research projects that are aimed at improving or increasing managerial efficiency and control, where maximum outputs are achieved through minimum inputs are rejected by critical researchers (Howcroft and Trauth, 2004:197).

- Critique of technological determinism: The idea that technological development follows its own rules and people and societies must adapt to technology and the pace that technology follows is challenged by critical researchers. Critical researchers point out that if a person has such a view, it allows the person who has interest in technology to increase power over other people. Critical researchers argue that people and society can shape the technology that is being developed (Howcroft and Trauth, 2004:197; Oates, 2006:298).

- Reflexivity: Critical researchers question the possibility of objective, value-free knowledge. Critical researchers are of the opinion that research projects and areas of development and knowledge are shaped by people who have power and vested interests. Critical researchers reflect on how their methods and they have influenced the knowledge outcomes emerging from research that has been done by them. Critical researchers are also aware that they are influenced by societal and organizational factors and history (Oates, 2006:298).

Willis (2007:392) describes the characteristics of critical research as follows:
- Nature of reality: The nature of reality is seen as material and external to the human mind.

- Purpose of research: The purpose of critical research is to discover local instances of universal power relationships and to empower the oppressed participants.
- Acceptable methods and data: Qualitative data is used more generally in critical research.
- Meaning of data: The data is used to enlighten and set free or emancipate the oppressed and is interpreted through ideology.
- Relationship of research to practice: The activities that take place in research, as well as practice are integrated and research will provide guidelines for practice.

In conclusion, it can be said that critical research goes beyond understanding a situation and challenges power structures which may dominate the situation or a person. Critical researchers assume that social reality is historically constituted and that it is produced and reproduced by people (Myers, 1997:242). Critical research focuses mainly on the disapprovals and conflicts in the current society, while trying to eliminate the objects or situations that lead to alienation and domination (Myers, 1997:242).

The main critique that has an influence on critical research is the fairness and authenticity of the research. The methods used for conducting critical research may differ from researcher to researcher, but the main method used is action research. Action research aims to contribute to both the practical concerns of the oppressed and the goals of social science. This is done by collaborating the practical concerns and the goals within an acceptable ethical framework (Myers, 1997:241).

The main reason why the critical researcher may choose to gather qualitative data is the difference between the natural world and a human's ability to communicate (Myers, 1997:241). This will allow the critical researcher to be able to help empower and overcome obstacles for participants of a study. The aim of critical research is to deliver critique and change within an oppressed reality. A comparison between the positivistic-, interpretivistic and critical research paradigm follows.

4.2.4 Comparison between research paradigms

In this section, a comparison between the discussed research paradigms (§4.2.1 to §4.2.3) is illustrated in Table 4.1 on page 79. The research paradigm's ontological beliefs, epistemological beliefs, the methodology that is followed, the unique characteristics, criticisms used for judging the quality of the research paradigm, the research methods, data collection methods, and the data analysis methods are compared.

From the information in Table 4.1 it is clear that the positivistic, interpretivistic and critical research paradigms each have unique and different ontological beliefs, epistemological beliefs, methodologies, characteristics and criticisms. Similarities between these three research paradigms can be emphasized in the research methods, data-collection methods and data-analysis methods. Research methods followed in both the positivistic and interpretivistic research paradigms include case studies, action research, as well as design and creation. Action research and design and creation are also used as methods of research in critical research.

Interpretivistic and critical research do not rely on collecting only quantitative data, as in positivistic research, but also qualitative data. Although the previous statement focuses on the similarity of the type of data to be collected, the data-collection methods for critical research differ from that of interpretivistic research. Questionnaires and interviews are data-collection methods used by the positivistic and interpretivistic research paradigms. These data-collection methods are discussed in §4.4 of this chapter. Computer-aided analysis techniques are used as a data-analysis method in the positivistic, interpretivistic and critical research methods. Coding is used as another data-analysis method in the interpretivistic and critical research paradigms, whereas the positivistic research paradigm rather depends on statistical techniques and visual representation of data through tables, charts and graphs.

Table 4.1 - Comparison between different research paradigms

	Positivistic Research	*Interpretivistic Research*	*Critical Research*
Ontology	Naïve realism - The reality is seen as real but can be apprehended.	Relativism – Local and specific constructed realities.	Historical realism – Social, political, cultural, economic, ethnic and gender values shape the reality known in a society or by a person.
Epistemology	Dualistic/Objectivist; findings true.	Transactional/Subjectivist; findings are created.	Transactional/Subjectivist; findings are value reconciled.
Methodology	Experimental and manipulative; verification of hypotheses.	Hermeneutical/Dialectical	Dialogical/Dialectical
Characteristics	Characteristics include: • the world exists independently of humans; • measurement & modelling; • objectivity; • hypothesis testing; • quantitative data analysis; and • universal laws.	Characteristics include: • multiple subjective realities; • dynamic, socially constructed meaning; • researcher reflexivity; • study of people in their natural social setting; • qualitative data analysis; and • multiple interpretations.	Characteristics include: • emancipation; • critique of tradition; • non-performative intent; • critique of technological determinism; and • reflexivity.
Criticisms, used for judging quality of the research paradigm	Criticisms include: • objectivity; • reliability; • internal validity; and • external validity.	Criticisms include: • credibility; • dependability & confirmability; • transferability; and • subjectivity vs excessive personal involvement.	Criticisms include: • fairness, • ontological authenticity; • educational authenticity; • catalytic authenticity; and • tactical authenticity.
Research methods	Research methods include: • experiments; • surveys; • case studies; • action research; and • design & creation.	Research methods include: • ethnography; • case studies; • action research; and • design & creation.	Research methods include: • action research; and • design & creation.
Data collection methods	Mostly quantitative data is collected by using: • questionnaires, and • interviews.	Both qualitative and quantitative data can be collected, but qualitative data is preferred: • interviews, • observations, • questionnaires, • documents, and • field notes.	Both qualitative and quantitative data can be collected, but qualitative data is preferred: • surveys, • case studies, • field notes, and • ethnographies.
Data analysis methods	Data analysis methods include: • tables, charts and graphs, • computer-aided analysis techniques; and • statistical techniques.	Data analysis methods include: • computer-aided analysis techniques; and • coding.	Data analysis methods include: • computer-aided analysis techniques; and • coding.

4.2.5 Research paradigm used for this study

The objective of this section is to depict the research paradigm employed in this study. This study attempts to answer the research question: Which techniques can be used by a company to ensure that social media is utilized effectively? In this study, the influence that social media can have on customer relationship management, marketing and branding, trend analysis and sales prediction is also investigated. A positivistic paradigm is adopted and best suited for this study as the researcher should keep an objective view of the reality.

The positivistic research paradigm was chosen for this study. Positivist research is the dominant research paradigm used when studying information technology in organizations (Orlikowski & Baroudi, 1991:6). Reviewing §4.2.1, it is clear that the positivist world view asserts that a theory is true only if it is repeatedly not proven to be false (Orlikowski & Baroudi, 1991:8).

The researcher assumes an objective physical and social world exists, independent of humans. The nature of this world can be apprehended, characterized and measured. The role of the researcher in this study is to discover this objective physical and social world by creating precise measures that will distinguish the dimensions of this world/reality in which the researcher is interested. To be able to create the measures, the researcher should firstly understand the dimensions. This can be done by studying literature and creating surveys in which one-to-one correspondence takes part between the researcher's model and the features within this world/reality (Orlikowski & Baroudi, 1991:9). For this study, different aspects of social media can be measured that can have an effect on customer relationship management and a company, whether to the benefit or detriment of the company.

The quality of this positivistic study is assessed according to the following criteria:
- Objectivity: The researcher's personal values and beliefs will not affect the investigation and a neutral and objective view will be adapted.
- Reliability: The research instrument consists of the questionnaire and techniques used to gather tweets. The questionnaire consists of a set of questions that is neutral, accurate and reliable. A few of the questions were adapted from previous studies as listed in section 1.3. The use of mathematical modelling and computer-aided analysis is used for analysing the questionnaires and the tweets.
- Internal validity: The research method, data gathering and data analysis were planned to capture the data relevant to this study and from appropriate sources.
- External validity: The questionnaires were sent to different people in different roles in different business areas. The Twitter data was gathered in such a way that the researcher

and participants did not have an effect on the process involved to analyse the data, and could not influence the information retrieved from the data analysis.

The characteristics of this study align with the characteristics of the positivist research paradigm, and the following are reasons for choosing the positivist research paradigm:

- Positivist research is regarded as the dominant paradigm when researching information systems, computer science or information technology (Oates, 2006:288; Orlikowski & Baroudi, 1991:6).
- Seeking to investigate the role that social media currently plays in companies and which techniques can be used by companies to utilize social media to their benefit
- The researcher is neutral and objective while analysing results and in terms of outcome (Oates, 2006:286).
- Using social media data, for example tweets, to improve customer relationship management by analysing the sentiment and opinion
- Statistical analysis and mathematical modelling are used when analysing Twitter data and data collected from surveys. The positivist research paradigm is associated with quantitative analysis which makes use of statistical analysis and mathematical modelling (Oates, 2006:38). The researcher can provide evidence of quantitative measures of variables. Patterns, relationships and behaviour need to be discovered from the data.

Interpretive and critical research have been gaining more interest when studying information systems, computer science and information technology. Researchers following these research paradigms are known to be more subjective and tend to make use of qualitative research methods (Lee, 1991:342). In the next section, the research methods associated with the positivistic paradigm are discussed, as well as the research method chosen for this study.

4.3 Research method

A research method can be defined as the way that data is collected (Harvey, 1990:1). No method of data collection can inherently be connected to the positivistic, interpretivistic or critical social research methodology but the research methods discussed in section 4.3.1 are most commonly associated with the positivistic research paradigm. The research methods discussed include: experiments, surveys, case studies, action research, design and creation. Surveys are used in this study and are discussed in section 4.3.2.

4.3.1 Research methods associated with the positivistic research paradigm

According to Oates (2006:299), the positivistic paradigm utilizes the following methods:

- <u>Experiments</u>

To test hypotheses and establish cause and effect, experiments are used (Oates, 2006:299). Experiments have been used by researchers for a long time and are the strategy of choice in the scientific method which is based on the positivist paradigm. Experiments tend to demand objectivity, and can be designed in such a way that the researcher(s) will not have an influence on the outcome of the experiment. Quantitative data and mathematical techniques are used to measure the inputs, outputs and the relationships between the inputs and outputs of an experiment.

Experiments aim for generalization and are therefore repeated, and a regularly occurring pattern, law or theory is then established. The expectation is formed that a regular occurring pattern or law does exist and that this should be found and demonstrated by the researcher(s). There are different experimental situations, which include laboratory experiments, field experiments and natural experiments (Mukherji & Albon, 2010:17).

- <u>Surveys</u>

Surveys are strongly associated with the positivistic paradigm (Oates, 2006:299). Surveys also tend to seek patterns, but it is important for the researcher to realize that when a survey is done, there is an assumption that patterns exist. The researcher should carefully decide on the sampling frame and sampling technique. This will ensure that the chosen sample is representative of a wider population. Quantitative data and statistical analysis are used in surveys to show patterns and relationships within the data, which is provided by the survey (Oates, 2006:299).

Surveys are sometimes designed in such a way as to be written up as a series of hypotheses. These hypotheses are then tested against data, but a survey cannot confirm the cause and effect, which an experiment can thus confirm. It must be kept in mind that a survey mainly confirms association. A survey can also be used in an interpretive way, which enables the researcher to establish a wider range of opinions about a particular subject (Oates, 2006:300).

- <u>Case studies</u>

There are researchers who argue that case study research can form part of a method used for positivistic research (Oates, 2006:301). A case study can be used to prove or disprove theories, similar to positivistic research. A researcher can choose a case so that the case is typical of a wider population, which then allows for general patters to be identified. Other characteristics of

positivism that this method supports are repeatability and generalization; this is because in some cases the outcome of a case study is a theory, which can then be applied to similar case studies.

- Action research

Action research can sometimes have a positivistic foundation. By testing existing theories, it can be determined whether the theories' predictions do occur in the research setting (Oates, 2006:301). If the researcher makes use of a control group the differences in outcomes can be determined and examined. Repeatability and generalization are also supported in this method, because action research can lead to one theory, which can then be tested in further studies. One characteristic, which is always questioned by positivistic researchers, is if this method could be used with objectivity. When a researcher makes use of action research, the researcher is an active participant, which causes the researcher not to act as an objective observer. Oates (2006:301) is of the opinion that any findings may be attributable to the researcher rather than any of the actions that were taken.

- Design and creation

A great part of design research in Information Systems and computing is based on the positivistic paradigm (Oates, 2006:302). The "cause leading to effect" characteristic comes to the fore because an assumption is made that computer systems are the means to specific ends. This means that the requirements of a system can clearly be stated and the designer is an objective outsider. Models that are produced by the designer are then seen as "true" representations of reality.

4.3.2 Research method used for this study

This study employs the survey research method. In this section, the term survey is defined and the process of planning and designing a survey is discussed. Why surveys were chosen as a research method for this study is also discussed in greater detail.

4.3.2.1 Defining surveys

A survey can be defined as a method to gather the same kind of data from a large group of people, by using a standardized and systematic way (Oates, 2006:93). The researcher has to identify patterns that can be generalized to a larger population than the targeted group (Oates, 2006:93). Creswell (2014:155) defines the term survey as a quantitative or numeric description of trends, attitudes or opinions of a population by studying a sample of that population. The definition of a survey that was used during this study is as follow: A survey can be defined as a method that creates the opportunity for a researcher to collect quantitative data, to identify

trends and generalizations within the sample of a population and to apply this newly gained knowledge to the rest of the population. Subsequently, the components that need to be considered during the planning and design of a survey are discussed.

4.3.2.2   The survey design process

A survey should be designed according to the purpose of the study and the rationale behind selecting a survey as research method (Creswell, 2014:157). Before choosing surveys as a research method, the researcher evaluated the advantages and disadvantages of the research method.

Oates (2006:104) describes the following advantages of surveys which are also applicable to this study and contributed to the researcher's choice of surveys as research method:

- Surveys provide a large coverage of people and the results collected can lead the researcher to draw generalized conclusions, because a wider population is represented.
- A vast amount of data can be produced in a short time and can save the researcher costs and time.
- Quantitative data analysis can be done; some researchers are of the opinion that quantitative data is the only suitable form of data.
- The scientific method has a positivistic philosophy and replication of studies is an important part of the scientific method. Data gathering and data analysis should be easy to replicate not only for other researchers, but also if the researcher wants to perform further tests. Surveys create the opportunity for the researcher(s) to replicate the data gathering process and to collect the same data from another sample.
- Surveys can be sent out by using the Internet (web questionnaires). This makes it easier for people who do not have good interpersonal or communication skills, and it is convenient for participants. In a survey no interviewer is present and the participant may be more willing to share information.

Oates (2006:105) also describes the following disadvantages of surveys. The researcher acknowledged the mentioned disadvantages.

- The content of surveys does not always include much detail regarding the topics being researched and the researcher may tend to focus more on the span of the coverage.
- Data gathered from surveys are subject to statistical analysis, meaning that the data gathered can be counted and measured. There are certain aspects of a research topic that cannot necessarily be translated into numbers and might therefore be overlooked.
- On-going processes, activities and change are not always focused on and surveys enable the researcher to gather data only within a certain time frame.

- Only associations can be made regarding the data gathered and the effect of a cause cannot be determined as with experiments.
- The accuracy of data gathered form Internet surveys (web questionnaires) cannot be judged, as the researcher cannot observe the participants body language and evaluate if the participant understands each question and the format of the answer that should be given.

After the advantages and disadvantages of the survey research method are evaluated, the design process of surveys should be planned. Oates (2006:94) breaks the planning and designing process of surveys into six different activities: data requirements, data-generation method, sampling frame, sampling techniques, response rate and non-responses and lastly, sample size. Creswell (2014:158) lists the following activities that should be included in the process of planning and designing a survey: population and sample, instrumentation, variables in the study, and the final activity, data analysis and interpretation. The following activities used to design the survey of this study are a compilation of the activities discussed by Oates (2006:94) and Creswell (2014:158):

- Activity 1 – identifying data requirements: The researcher should decide what data should be generated. The data that need to be generated for this study are directly associated with the research question and some of the data requirements are indirectly related to the research question.
- Activity 2 – identifying variables in the study: Identified variables are connected to each study objective as listed in §1.4. In Table 4.2 cross-references are made between the variables, questions and study objectives.
- Activity 3 – data-generation method/instrumentation: Data will be generated by distributing questionnaires via the Internet (§4.4.1). The data generated from this method will be used to answer the research question. Twitter data (tweets) will also be gathered and sentiment analysis and opinion mining will be done on this data. The process of data generation regarding Twitter data will be discussed in Chapter 6.
- Activity 4 – identifying a population and sample: By performing this activity, the researcher will identify the population, the size of the population, the sampling design and selection process/sampling technique. The details of this activity will be elaborated on in §4.3.2.3. The researcher respected the participants' expectation of anonymity and confidentiality. The rights of the participants were made clear: a participant did not have to participate, could withdraw at any time and gave informed consent. The researcher did not intimidate people to participate in the survey and did not deceive people regarding the research being done.
- Activity 5 – deciding on the sample size and response rate: By giving a sound explanation of the purpose of the study and the survey, the researcher attempted to persuade more people

to participate. The researcher kept reminding people and politely asking people to participate in the survey. By identifying the characteristics of non-responses, the researcher noticed that non-responses came from companies not using social media. The researcher took into account the none-response rate of participants when deciding on the sample size.

- Activity 6 – data analysis and interpretation: Statistical analysis will be done on data gathered from the questionnaires, statistical methods applied and results will be discussed in Chapter 5. Data-mining techniques will be applied on Twitter data by using open-source software. The process and results will be discussed in Chapter 6.

Next, the application of the survey research method in this study is discussed.

4.3.2.3   Application of surveys in this study

A survey was conducted to identify which techniques, strategies or frameworks are currently used to manage social media of a company. An anonymous survey was conducted to ensure that an employee's opinion regarding the utilization of social media within the company could be captured. The target participants were preferably employees who had knowledge regarding the use of social media by the company. Only one survey was done for a company. If the company consisted of more than one department, the survey was done for each department in the company. This also contributed to ensuring that unique data could be gathered and that the same data was not gathered continuously.

Participating companies were formally addressed to determine whether they would be willing to participate in the study. An e-mail with a cover page and hyperlink to the questionnaire was then sent to the participating company. Anonymity of the participant was ensured as the completed survey's answers were gathered with no reference or any personal identification of the participant. A Google Form was used for the survey which stored the answers in a sheet on Google Drive. The cover page and hyperlink were also distributed on social media research groups on LinkedIn. The same e-mail that was sent to companies was also sent to former students of North-West University, Potchefstroom campus. A total of 122 questionnaires were completed correctly.

4.4   Data-generation method

Data-collection and data-analysis techniques are important to the researcher to ensure that the findings of a project can be represented precisely. According to Oates (2006:245), quantitative data is primarily used and analysed by positivistic researchers. There is a wide range of established techniques that can be used for analysing quantitative data.

- Questionnaires used for surveys

A questionnaire can be defined as a set of questions, which is assembled in a pre-determined order and that provides the researcher with data, which can be analysed and interpreted

(Oates, 2006:219). Questionnaires are used in the survey research strategy. By analysing all the responses, patterns and generalizations can be found regarding the population that participated in the questionnaire. Questionnaires can also be used in other methods used in the positivistic paradigm, for instance case studies, action research and even in design and creation (Oates, 2006:219).

Questionnaires can either be self-administered or researcher-administered (Oates, 2006:219). If a questionnaire is self-administered, then the participant will complete the questionnaire without the researcher being present. If the questionnaire is researcher-administrated, then the researcher will ask the participant every question, and then the researcher will write down the response.

- Interviews
According to Oates (2006:36), an interview is a particular kind of conversation, where the researcher has control over the agenda and most of the questions asked for the duration of the interview. Interviews can either be done face-to-face, telephonically or by making use of computer-assisted personal interviewing.

4.4.1 Data-generation method used for this study
The data-generation method chosen to complement the survey research method was web-based questionnaires. Questionnaires are predominantly used to gather quantitative data. Quantitative research is about explaining phenomena by collecting quantitative data, which are mainly analyzed by mathematical techniques (Balnaves & Caputi, 2010:9). Questionnaires were decided on because it is best suited to the situation in which the study takes place.

A questionnaire is best suited if the researcher wants to obtain data from a large population (Oates, 2006:220). It is important for the researcher to standardise the data and to ensure that the identical questions are asked and answered by each participant (Oates, 2006:220). The costs involved with questionnaires should also be kept in mind by the researcher. The aim of designing the questionnaire is to gather reliable data at a reasonable cost.

The researcher considered the advantages and disadvantages of using a questionnaire as data-generation method. The following list of advantages was compiled by the researcher and gathered from the listed sources (Jones *et al.,* 2008:16; Oates, 2006:229):
- Relative low costs of data collection and processing can lead to vast amounts of data collected. Larger numbers of the target population can be reached.

- Minimal training is required by the participant and the questionnaire prepared for this study is self-administrative. The use of pre-defined answers makes it easier for the participant to complete the questionnaire, which can then easily be analyzed by the researcher.

- Questionnaires can be delivered verbally, by telephone, electronically as e-mail attachments or as a web link, which will then lead the participant to the questionnaire.

- When using a web questionnaire, the participant will not experience a feeling of being at a distance from the researcher, because on-line instructional pages are not limited to text and can be provided to the participant.

- Data entry on web-based questionnaires can be controlled through real-time error checking and correction, ensuring that the participant is guided through the process. By performing this control, the researcher can ensure that the completed questionnaire is submitted.

The following list of disadvantages was compiled by the researcher and gathered from the listed sources (Jones *et al.,* 2008:16; Oates, 2006:229):

- The honesty of answers cannot be checked by the researcher, especially the answers to open questions in the questionnaire.

- Little or no contact between the researcher and the participant(s) can lead to low response rates.

- The participant can experience frustration with pre-defined answers and may refuse to answer, or an associated bias toward a question(s) can develop.

- Converting a hard copy questionnaire to a web-based questionnaire requires more effort from and expertise by the researcher.

- Delivering a web-based questionnaire automatically excludes possible participants who do not have access to a computer or an Internet connection.

- Self-administered questionnaires are inappropriate for participants who are visually handicapped or have poor literacy skills.

Subsequently, the application of questionnaires as data-generation method for this study is discussed.

4.4.1.1 Application of questionnaires for this study

The questionnaire was designed based on related literature gathered. The questionnaire was pre-tested and approved by Statistical Consultation Services, North-West University, Potchefstroom campus.

The steps taken to collect data were as follows:

1. Identify companies that make use of social media for customer relationship management, branding, marketing, etc. and that are willing to participate in the study.

2. Send the cover letter and hyperlink to the questionnaire to the participating company or employee within a department. Contact information of the researcher was also provided if the employee or a company had any questions regarding the study or the content of the questionnaire.

3. A completed questionnaire was returned and the results were saved directly into an Excel spreadsheet. The Excel spreadsheet was saved on a Google Drive account, which could only be accessed by the researcher.

4. Some of the participants communicated that they had completed the questionnaire and a word of appreciation was sent to these participants.

5. The cover page and hyperlink to the questionnaire were distributed on social media groups on LinkedIn numerous times.

6. A friendly follow-up was done regarding participants who assured the researcher that they would complete the questionnaire.

Even though the researcher followed the mentioned steps, challenges did arise during the distribution of the questionnaire:

- The average comment was that the questionnaire was too long and took too much time out of or interfered with the participant's daily tasks.

- There was a possibility that only a small number of questionnaires would be completely filled in and that not all sections would be completed.

- Even though a considerable number of e-mails was sent, only a small number of the questionnaires was completed.

- Companies and employees are not willing to participate in research done on social media, even though the company can only benefit from the research. This may be because a company is not willing to disclose the techniques, strategies and frameworks used for social media as they feel that another company might gain advantage from the information.

Even though the researcher experienced these challenges, it was a very enjoyable experience to learn more regarding the use of social media within companies today. Next, the design of the questions in the questionnaire will be discussed.

4.1.1.2   The design of the questions in the questionnaire

Each question was designed in such a way that it was relevant to the overall questionnaire and the purpose. The questions were developed to be unambiguous, specific and objective. Both open-ended questions and closed questions were used in the questionnaire. Open-ended questions allow the participant to decide on the answer that best suits them (Oates, 2006:222). Open-ended answers require coding by the researcher. This process will also be discussed in Chapter 5. Closed questions force the participants to choose from a range of answers that have been pre-defined by the researcher (Oates, 2006:222).

The questionnaire comprised a total of 47 questions and was divided into four sections:

Section A – General information

The researcher identified in which business area and type of enterprise the participant was employed. Different business areas and enterprises may adopt a different strategy when managing social media platforms. This section consisted of three questions. Two of the questions were multiple-choice to identify in which business area the participant was currently working and the number of employees employed by the company. The third question was a Likert-scale question ranging from not useful (1) to very useful (4), which determined the participant's opinion regarding the use of social media by the company.

Section B – Social media platforms

The aim of this section was to investigate the effectiveness of different social media platforms and to identify which social media platforms are mostly used by companies. In this section, the researcher also investigated the awareness of customers and clients regarding social media platforms from the participants' and company's perspectives. In this section, data was gathered regarding the employee(s) managing a company's social media platforms, whether the employee(s) had special training, was appointed in-house or whether the company was outsourcing the management of its social media platforms.

This section consisted of a total of 20 questions, and included six open-ended questions and 14 closed questions. Of the 14 closed questions; six questions were dichotomous (yes or no), two questions were multiple-choice, and six questions could be answered according to a specified Likert scale (1 – 4). Of the six Likert-scale questions, three questions tested the effectiveness of a list of social media platforms (not effective (1), moderately effective (2), effective (3) and extremely effective (4)). One of the scale questions captured the opinion of the participant regarding different uses of social media by the company and the participant was asked to rate a list of uses on a scale from 1 (not used) to 4 (almost always used). Two of the scale questions

allowed the participant to express his/her level of agreement or disagreement where 1 was coded as strongly disagree, and 4 was coded as strongly agree.

Section C – Social media strategy

The aim of this section was to identify whether the company's social media goals are similar to the business objectives. In this section, the researcher gathered data regarding techniques, strategies and frameworks currently being used by the company. Reasons why companies are not using social media platforms more often are explored, as well as the steps that a company would take if a social media crisis happened.

This section consisted of a total of 14 questions, and included seven open-ended questions and seven closed questions. Of the seven closed questions, three questions were Likert-scale questions and four were dichotomous questions (yes or no). Two of the scale questions tested the integration of social media goals regarding the business objectives (where 1 indicated - not integrated and 4 indicated - fully integrated) and the effectiveness of a strategy or social media framework specified by the participant (where 1 indicated - not effective and 4 indicated - very effective). One of the scale questions allowed the participant to express his/her level of agreement or disagreement regarding a list of options on why companies do not use social media platforms more often where 1 was coded as strongly disagree and 4 was coded as strongly agree.

Section D – Social media metrics

The aim of this section was to identify which metrics companies are currently using to track social media efforts. The most popular content posted on social media platforms was also identified. The researcher also identified which techniques the company was using to pull analytics and data from Twitter and Facebook. This section consisted of a total of 10 questions, which included five open-ended questions and five closed questions. Of the five closed questions, two were dichotomous questions (yes or no), one question was multiple-choice and two questions were Likert-scale questions. The multiple-choice question gathered data regarding how often content was posted by the company on social media platforms. One of the scale questions gathered data regarding the effectiveness of the type of data based on a scale from 1 (not effective) to 4 (extremely effective). The level of agreement and disagreement of the participant was also tested regarding linking a company's Twitter posts to the company's Facebook page.

Table 4.2 depicts the design of the questionnaire in more detail and contains the following information:

- Question,

- reason for the question,

- to which research objective data gathered from the question will contribute,

- the research variable associated with the question,

- the question type, and

- a reference to literature is given that led to the development of the question.

Questionnaires were not the only method used for data generation and the process of gathering Twitter data is discussed in §4.4.2.

Table 4.2 - Design of the questionnaire for this study

Section	Question	Reason for question	Question contribution towards research objectives	Research variable	Question type	Reference (Optional)
Section A - General information	**1.** In which of the following business areas are you mainly employed in?	To determine if the responses given throughout the questionnaire differs according to business area.	SO8	Business_area	The type of business area will be determined through the use of a multiple choice question: - Retail, sales & marketing - Banking, finance & accounting - Insurance - Information technology & communications - Legal - Government	- Petre (2013:23) - §3.2.1. in Chapter 3 of this dissertation.
	2. Roughly how many full-time employees are currently working for the company?	Determining if the size of the company has an impact on how the company utilizes social media.	-	Size	The number of employees will be determined through multiple choice: - 1 – 9 (micro-enterprise) - 10 – 49 (small enterprise) - 50 – 249 (medium-sized enterprise) - 250+ (large enterprise)	- Crews and Stitt-Gohdes (2012:2)
	3. How useful do you think social media has been for the company?	The respondent's opinion regarding the usefulness of social media use within the company is determined.	SO9	Usefulness	The level of agreement will be determined by allowing the respondent to choose from a scale ranging from not useful to very useful.	-
Section B - Social media platforms	**1.** Are you allowed to visit external social media platforms, e.g. Facebook, Twitter, LinkedIn, etc. during working hours?	Link can be established with question 2 in Section B to determine whether training was done on using social media platforms that the company's reputation would not be damaged.	SO8	Visits	The number of employees that are allowed to visit social media platforms during working hours will be determined by giving dichotomous answer (yes or no).	- Lardi and Fuchs (2013:23) - §2.3. and Table 2.3. in Chapter 2 of this dissertation.
	2. Does the company have social media training across the company to ensure that employees understand how to engage on social media platforms and consistently represent the company?	Determine if employees have been made aware that there are legal liabilities and other aspects that should be adhered to when using social media platforms and using the company name etc. in for example a post or tweet.	SO9	Experience/Training	By using a dichotomous question (yes or no) there will be determined if employees are trained to use social media platforms in a way that it won't be harmful to the company's reputation.	- Merril *et al.* (2011) - Rideout (2014:133) - §2.3.7. in Chapter 2 of this dissertation.

Section	Question	Reason for question	Question contribution towards research objectives	Research variable	Question type	Reference (Optional)
Section B - Social media platforms	**3.** How long has the company been using social media platforms?	This question will contribute to determining the number of months or years that the company has been using social media platforms as well as the company's experience with social media platforms. This question will also determine if the length of time corresponds to other factors.	SO8 & SO9	Experience/Time_period	The time period will be determined through multiple choice: - 1 – 3 months - 4 – 8 months - 9 – 12 months - 1 – 2 years - 3 – 4 years - 5+ years	-
	4.1. Please rate the effectiveness of the following social media platforms used for marketing by the company? (where applicable)	Determining the most popular social media platforms currently being used by companies/businesses for marketing.	SO8	Social_media_platform	The following social media platforms will be rated on a scale from not effective to extremely effective: - Facebook - Twitter - LinkedIn - Google+ - YouTube - Pinterest - Instagram - Tumblr	- Geho (2010) - Lardi and Fuchs (2013:23) - Paniagua and Sapena (2014:723)
	4.2. Please rate the effectiveness of the following social media platforms used for branding by the company? (where applicable)	Determining the most popular social media platforms currently being used by companies/businesses for branding.	SO8	Social_media_platform	The following social media platforms will be rated on a scale from not effective to extremely effective: - Facebook - Twitter - LinkedIn - Google+ - YouTube - Pinterest - Instagram - Tumblr	- Geho (2010) - Lardi and Fuchs (2013:23) - Paniagua and Sapena (2014:723)
	4.3. Please rate the effectiveness of the following social media platforms used for customer relationship management by the company? (where applicable)	Determining the most popular social media platforms currently being used by companies/businesses for customer relationship management.	SO8	Social_media_platform	The following social media platforms will be rated on a scale from not effective to extremely effective: - Facebook - Twitter - LinkedIn - Google+ - YouTube - Pinterest	- Geho (2010) - Lardi and Fuchs (2013:23) - Paniagua and Sapena (2014:723)

Section	Question	Reason for question	Question contribution towards research objectives	Research variable	Question type	Reference (Optional)
					- Instagram - Tumblr	
	4.4. Are there any other social media platforms used by the company that is not listed in the previous questions?	Determining if there are any other social media platforms not listed in question 4.1. – 4.3.	SO8	Social_media_platform	An open-ended answer will be required from the respondent.	-
	5.1. How does the company determine awareness of customers and clients regarding social media platforms being used?	This question will examine the methods that companies are using to determine customer and client awareness.	SO6	Customer_awareness	An open-ended answer will be required from the respondent.	-
	5.2. According to the company's perspective, rate how aware customers and clients are regarding social media platforms being used.	This question will determine the company's perspective regarding how aware customers and clients are.	SO6	Awareness	The level of awareness will be determined by using a scale that can be rated from not aware to extremely aware.	-
Section B - Social media platforms	**6.1.** According to your opinion rate each of the following uses of social media by the company.	This question will determine the main purpose(s) of the company's social media use.	SO8	Uses	Social media can be used by companies for different reasons and the use will be determined through a scale ranging from almost never used to almost always used: - To advertise products/services. - To offer promotional offers/items. - To increase brand awareness. - To gain feedback from customers. - To engage (in conversation) with customers. - For business to business purposes e.g. LinkedIn. - To analyse the competition in the industry. - Other, please specify…	- Table 2.3. in Chapter 2 of this dissertation.
	6.2. List any other uses of social media platforms which is not listed above.	This question will determine the main purpose(s) of the company's social media use.	SO8	Uses	An open-ended answer will be required from the respondent.	-
	7. Indicate your level of agreement regarding the following statements: - Social media is effective for brand awareness. - Social media has revolutionised the marketing of a product/service. - Social media is effective for selling a product/service.	This question will test the perceptions regarding the effectiveness of social media platforms.	SO8 & SO9	Agreement_level	The agreement level will be determined through a rating scale ranging from strongly disagree to strongly agree.	- Geho (2010) - Lardi and Fuchs (2013:23) - Paniagua and Sapena (2014:723)

Section	Question	Reason for question	Question contribution towards research objectives	Research variable	Question type	Reference (Optional)
	- Social media is effective for advertising a product/service. - Social media is great or engaging with customers.					
	8. Indicate your level of agreement with the following statement: This social media platform allows a customer to express sentiment and his/her opinion. - Facebook - Twitter - LinkedIn - Google+ - YouTube - Pinterest - Instagram - Tumblr	Determines the most popular social media platforms used for customer relationship management. This question can be linked to question 4.1 to 4.4 of Section B.	SO6	Customer_relationship_growth	The agreement level will be determined through a rating scale ranging from strongly disagree to strongly agree.	- Lardi and Fuchs (2013:23) - Rygielski *et al.* (2002:491)
Section B - Social media platforms	**9.** Is your company paying to promote posts on social media platforms?	The researcher would determine if a company is willing to pay to promote their product(s)/service(s) on social media platforms.	SO8	Cost	The promotion of posts will be determined through a dichotomous answer (yes or no).	-
	10.1. According to your opinion does different social media platforms attract a different audience?	This question would determine if different social media platforms would attract a different audience.	SO6 & SO8	Client/Customer_type	There will be determined if different social media attract different audiences through a dichotomous question (yes or no).	- Wiegand (2015)
	10.2. How does the company determine the audience of their social media platforms?	Identifies methods that a company uses to determine the audience for which social media content should be adapted for.	-	Client/Customer_type	An open-ended answer will be required from the respondent.	-
	11. What is a limitation that the company has experienced on a social media platform and how did the company overcome this?	Determines any limitations that can be listed regarding to social media platforms as well as possible solutions.	SO8 & SO9	Limitations	An open-ended answer will be required from the respondent.	- §2.3.7. in Chapter 2 of this dissertation.
	12. 1. Has the company hired or appointed employees specifically to manage the company's social media platforms?	This question will determine if companies are prepared to hire employees or to outsource the management of social media platforms.	SO8	Strategy	The respondent will choose from a dichotomous answer (yes or no).	- Erdoğmuş and Çiçek (2012:1355)
	12.2. If you have answered Yes to the question 12.1. have the employees appointed to manage the social media platforms received training or own a certain qualification, if Yes please specify.	Determines if employees need to have additional training and expertise in managing social media platforms.		Qualified/Expertise	An open-ended answer will be required from the respondent.	- Erdoğmuş and Çiçek (2012:1355)

Section	Question	Reason for question	Question contribution towards research objectives	Research variable	Question type	Reference (Optional)
	12.3. Are these employee(s) appointed in-house or outsourced?	This question will determine what the trend is regarding social media experts, if they are appointed in-house or outsourced.		In-house/Outsourced	The respondent will choose from a dichotomous answer (yes or no).	-
Section C - Social media strategy	**1.** What are the goals that the company want to achieve by using social media platforms?	This question will help determine if companies in the same business area have the same social media goals (Section A, question 1)	SO8	Goals	An open-ended answer will be required from the respondent.	- He *et al.* (2013:464) - Jeffrey (2013:2) - Kaplan and Haenlein (2010)
	2. According to your opinion rate how integrated the company's social media goals are with the overall business objectives.	Determines the employee's opinion regarding the integration of the company's social media goals with the business objectives.	SO8	Integration_of_goals	The respondent will have to rate this question according to a scale ranging from not integrated to fully integrated.	- He *et al.* (2013:464) - Jeffrey (2013:2) - Kaplan and Haenlein (2010)
	3. What are the steps that the company follows to encourage social media followers to become customers/clients?	Determines what methods are used to ensure that a follower can become a potential customer/client.	SO6 & SO8	Growth	An open-ended answer will be required from the respondent.	-
	4.1. Does the company make use of a framework or strategy when using social media platforms?	This question will determine if the company is currently following a social media strategy or making use of a social media framework to act as guidance when using social media platforms.	SO8	Use_of_framework	The respondent will choose from a dichotomous answer (yes or no).	- AMEC (2013) - Jeffrey (2013) - Paine (2015)
	4.2.1. If you have answered Yes to question 4.1. please name the framework or strategy being used.	Determines which frameworks are currently being used.	SO8	Framework/Strategy	An open-ended answer will be required from the respondent.	-
	4.2.2. Rate the effectiveness of the named framework or strategy.	Determines the effectiveness of the named framework.	SO8	Effectiveness_framework/strategy	The respondent will have to rate the named framework on scale ranging from not effective to very effective.	-
	5. Does your company make use of a dashboard to manage the different social media platforms at the same time?	This question will determine whether the company manages each social media platform being used separately or if the company is using other software to manage each of the social media platforms at the same time.	SO4 & SO8	Customer_relationship_growth	The respondent will choose from a dichotomous answer (yes or no).	-
	6. How is your company monitoring competitive companies, products or services?	Determines whether the company is keeping watch on their competitors within the same business area or industry (question 1, Section A).	SO8	Monitoring_of_competitors	An open-ended answer will be required from the respondent.	- Cvijikj and Michahelles (2011:6) - Golfarelli *et al.* (2004:1) - Rygielski *et al.* (2002:493)

Section	Question	Reason for question	Question contribution towards research objectives	Research variable	Question type	Reference (Optional)
Section C - Social media strategy	**7.** According to your opinion is there a lack of control on the information being dispersed about the company, when using a social media platform?	To determine if a company feels that there is a lack of control and trust when publishing content on social media platforms and whether it can be harmful to the company's reputation.	-	Control	The respondent will choose from a dichotomous answer (yes or no).	- §2.3.7. in Chapter 2 of this dissertation.
	8.1. According to your opinion should the company be using social media more in its day-to-day operations?	This question determines the opinion regarding the usefulness of social media according to an employee of a company/business.	SO8	Use_of_social_media	The respondent will choose from a dichotomous answer (yes or no).	-
	8.2. If you answered Yes to question 8.1. please state why you as employee feel that the company should be using social media more in its day-to-day operations.			Employee_opinion	An open-ended answer will be required from the respondent.	
	8.3. Please express your level of agreement regarding the following statements on why the company does not use social media platforms more often.	Determines why a company is not using social media as often as is expected of the company employees.		Factor_influencing_use	The respondent will rate the following options on a scale ranging from strongly disagree to strongly agree: - Not enough time. - Not enough resources e.g. staff, finances, etc. - Don't think that social media is useful to the organization. - Have insufficient knowledge on how to use social media so the company will benefit. - Other, please specify…	- Hill (2015) - Lee (2015) - Torr (2014)
	8.4. Please list any other reasons, than those listed above, why the company does not make use of social media platforms more often?	Determines what other reasons can contribute to why the company does not make use of social media platforms more often.	SO8	Factor_influencing_use	An open-ended answer will be required from the respondent.	-
	9. How does your company handle a social media crisis? For example if one of the social media platforms of the company is hacked.	Possible solutions will be listed regarding social media crises. The answers to this question can also be linked to question 11 of Section B.	SO8 & SO9	Crisis	An open-ended answer will be required from the respondent.	- §2.3.7. in Chapter 2 of this dissertation.
	1.1. Has the company established metrics to track social media efforts to the company's objectives and goals?	Determines whether the company is aware that social media efforts can be measured according to the social media goals as well as the company's objectives and goals.	SO5 & SO8	Metrics	The respondent will choose from a dichotomous answer (yes or no).	- AMEC (2013)
	1.2. If you have answered yes to question 1.1. what metrics are the company using to track social media efforts?	Identifies what metrics companies are using to track social media efforts.	SO5 & SO8	Metrics	An open-ended answer will be required from the respondent.	-

Section	Question	Reason for question	Question contribution towards research objectives	Research variable	Question type	Reference (Optional)
Section D - Social media metrics	**2.1.** If the company has a web site, is the company monitoring and gathering analytics and data from the web site?	This question will determine if the company is also gathering analytics and data not only from their social media platforms but also from their web site.	SO4, SO5 & SO6	Analytics	The respondent will choose from a dichotomous answer (yes or no).	- Al-Azmi (2013:2) - Cormode & Krishnamurthy (2008)
	2.2. If you have answered Yes to question 2.1. how is the company gathering analytics and data from the web site?				An open-ended answer will be required from the respondent.	-
	3.1. How often is content posted on social media platforms?	Determines how regular companies are posting new content on social media platforms.	SO8	Posting_of_content	The respondent will be able to choose from the following multiple choice options: - Multiple times per day - Daily - Weekly - Every fortnight - Monthly	-
	3.2. Please rate the effectiveness of the type of content posted on social media platforms (where applicable).	Determines the most popular type of content that is being posted by companies on social media platforms.	SO4, SO5 & SO8	Type_of_content	The respondent will have to rate the following according to a scale ranging from not effective to very effective. - Text - Images, photos, etc. - Video's	- §2.3. in Chapter 2 of this dissertation.
Section D - Social media metrics	**4.1.** According to you opinion do you agree that it is a good idea to link Twitter posts to automatically post to the company's Facebook page?	This question will determine whether employees of a company feel that Twitter posts should also appear and be converted to Facebook posts immediately.	SO4	Success_of_linking	The respondent will have to indicate the level of agreement on a scale ranging from strongly disagree to strongly agree.	-
	4.2. How does the company pull analytics and data from Twitter?	Determines which methods companies are currently using to pull analytics and data from Twitter and to determine whether there are any new methods that the researcher was not aware of and can be used for future research.	SO4 & SO5	Methods	An open-ended answer will be required from the respondent.	- Adedoyin-Olowe *et al.* (2013)
	4.3. How does the company pull analytics and data from Facebook?	Determines which methods companies are currently using to pull analytics and data from Facebook and to determine whether there are any new methods that the researcher was not aware of and can be used for future research.		Methods	An open-ended answer will be required from the respondent.	-
	5. How does the company measure social return on investment (ROI)?	This question will allow methods being used by companies to measure social return on investment to be listed and can be linked to question 1 of Section C.	SO8 & SO9	Customer_relationship_growth	An open-ended answer will be required from the respondent.	-

4.4.2 Data-generation method used for Twitter data

Twitter data (tweets) can capture information, and customers' feelings and opinions about any topic, product, brand or service imaginable. This is the main reason why the researcher chose to use tweets to perform sentiment analysis and opinion mining. Today, there is a variety of ways to gather Twitter data; a person can use a programming language, such as Python with a library known as Tweepy, Twitter Firehose, Google Sheets, etc. (Russel, 2014:9).

Whichever process a person chooses to gather Twitter data, the process will involve using Twitter's Streaming Application Programming Interface (API). An Application Programming Interface is a set of protocols, routines and tools that can be used to build software applications (Beal, 2016). Before gathering Twitter data, a person should have a Twitter account. When using a programming language, such as Python in conjunction with Twitter's API, a person will have to gain access to Twitter's API. In order to gain access to Twitter's API, a person will need an API key, an API secret, an access token and an access-token secret.

A person will thus have to create a new application in Twitter to obtain the API key, API secret, access token and access-token secret. The data format will depend on the method used when gathering Twitter data. For example, when using Python and the Twitter Streaming API, the data will be returned in JSON format. For this study, the researcher used Twitter's Streaming API in conjunction with Google Sheets, as this method allows for free, easy and fast gathering of Twitter data.

Google Sheets offers a simple add-on, known as Twitter Archiver. This add-on allows a person to search Twitter for a specific topic, keyword or hashtag and saves the collected data to a spreadsheet. This add-on gathers all the tweets it can from Twitter's history and updates the spreadsheet with new tweets every hour. If a person has a paid account tweets will be refreshed every 10-15 minutes. Twitter Archiver will gather the last 1 000 tweets or tweets that have been posted in the last 5-7 days.

The researcher created a new search rule, by using the add-on, which searched for specific keywords. For this study a company's name was searched for and tweets regarding the company were gathered. Other data that was also imported from Twitter included the tweet's retweet, the tweeter's follower's count, the application used to create the tweet and the location. Tweets were gathered from three companies: an international company in retail, and two South-African companies, one company in banking and the other company in insurance. This data was used to conduct sentiment analysis and opinion mining and the analysis of data and results are discussed in Chapter 6 of this dissertation.

4.5 Data analysis

The idea behind data analysis in the positivistic paradigm is to search for general patterns in the data and draw conclusions according to these patterns. One of the easiest ways of data analysis used is tables, charts or graphs, which enables the researcher to illustrate patterns found in the data (Oates, 2006:246). Visual aids for quantitative data analysis include:

- Tables: This visual aid can be used with all types of data, and can be developed by making use of word processing software (Oates, 2006:249).
- Bar charts: Frequencies can be displayed by making use of a bar chart.
- Pie charts: If the researcher wants to display data as proportions, a pie chart can be used. This can make it easier for the reader to read (Oates, 2006:252).
- Scatter graphs: If the researcher wants to illustrate the relationship between two variables, a scatter graph can be used (Oates, 2006:252). The data is plotted as points on a graph and the x-axis and y-axis represents the values of each of the variables separately.
- Line graphs: The researcher can use a line graph to represent the trends in data.

Descriptive statistical techniques, as well as complex statistical techniques are used to enable the researcher to establish whether the patterns found in the data does exist and are not a result of chance (Oates, 2006:246). The use of statistical and mathematical analysis agrees with the definition of positivistic research because these analysis techniques support generalization, which is one of the main characteristics of this research methodology. Statistical techniques include calculating the mean, the median, describing the range, etc. The relationships in data can also be found by calculating the correlation coefficients (Oates, 2006:258).

For evaluating quantitative data analysis, the researcher should keep in mind the type of data, for example if the data is nominal or ordinal and if the correct visual aids are used for illustrating the data. There must be a balance between the use of statistical methods to interpret the data and the researcher's own interpretations. The data analysis methods used in this study will subsequently be discussed.

4.5.1 Data analysis used in this study

Quantitative data analysis was used to analyze and evaluate the data collected by means of questionnaires. Twitter data collected (§4.4.2) was also analyzed by using statistical methods. The statistical methods, data- and text mining techniques used to analyze the tweets, will be discussed along with the process used to determine sentiment and opinions in Chapter 6. In this section the focus will be on the definition of quantitative data analysis and the reasons why the researcher chose quantitative data analysis to analyze the questionnaires.

4.5.1.1 Defining quantitative data

Oates (2006:245) defines quantitative data as data or evidence that are based on numbers. This type of data is normally generated if the researcher makes use of experiments and surveys. Creswell (2014:224) defines quantitative data analysis as descriptive and inferential numeric analysis. The definition of quantitative data that will be used for this study is: Data which is based on numbers that can be analyzed to identify patterns, relationships and behaviour and draw conclusions. The reasons for choosing quantitative data analysis and how quantitative data analysis was applied in this study are discussed next.

4.5.1.2 Reasons for choosing quantitative data analysis and application to this study

Quantitative data analysis was conducted using the Statistical Analysis System (SAS) and the validity of the data analysis was checked by Mrs. Wilma Breytenbach of Statistical Consultation Services (SCS), North-West University, Potchefstroom Campus. Before the data could be used, some of the data had to be coded into numbers in order to carry out quantitative data analysis with ease. Visual aids, such as tables, charts and graphs will be used to organize the results of the data. Statistical techniques will also be used to help identify patterns and draw generalized conclusions. Statistical techniques used for analysis will be discussed in more detail, along with the results of the questionnaire in Chapter 5.

The type of data should also be considered when applying analysis techniques (Oates, 2006:246). Data collected from the questionnaire could be divided into the following types:

- *Nominal data:* This type of data usually describes different categories (Oates, 2006:247). For example, in Section A of the questionnaire the different business area needs to be indicated by the participant. Each business area can be assigned to a different number and the frequency of each business area can then be determined.
- *Ordinal data:* Likert-scale questions were used to collect data and the most of the data collected from the questionnaire can be characterized as ordinal data. Ordinal data is the allocation of numbers to a quantitative scale (Oates, 2006:247).

Qualitative data (non-numeric data) was also captured from the questionnaire and the researcher used quantitative (numerical) analysis and theme analysis to analyze this data.

The advantages and disadvantages of quantitative data analysis must be considered before choosing to collect quantitative data. The following list of advantages was compiled from the listed sources (De Vault, 2016; Madrigal & McClain, 2012; Oates, 2006:263):

- The analysis is based on well-established statistical techniques, such as calculation of the mean, median, and standard deviation. Inferential statistical techniques can also be included, such as t-tests, ANOVA's and multiple regression correlations. The statistical

tests done can be checked, because the analysis is based on measured quantities and not subjective impressions.

- Quantitative research provides data that is descriptive and allows for a broader study.
- Some researchers believe that quantitative research is the only valid form of research, as scientific measurements can be recorded.
- Large volumes of data can be analyzed in a short time because statistical computer programs can be used.
- Quantitative analysis allows the researcher to make generalization regarding the population that has been investigated.

The disadvantages and weaknesses of quantitative research are given in the following list of disadvantages regarding quantitative research, which was compiled from the listed sources (De Vault, 2016; Madrigal & McClain, 2012; Oates, 2006:263):

- The researcher may lose sight of the purpose of the research being done, when using computers and statistical software to analyze the data without understanding the software properly.
- The researcher should indicate which statistical tests will be used and what kinds of quantitative data are required before data generation can be started.
- When analyzing the data, decisions taken by the researcher can influence the results for example choosing scales on the x- and y-axis.
- The reflection and opinions of participants regarding a specific topic cannot be gathered when closed questions are used for data gathering. The data gathered can then be perceived as a false representation.

After consideration of the advantages and disadvantages, quantitative research and data analysis were chosen because they are more reliable and objective, and more suited for working in the positivistic research paradigm. Mathematical modelling and statistics can be used to generalize a finding, an assumption is made that the sample is a representation of the population, and relationships between variables are identified.

4.6   Conclusion

The objective of this chapter has been to discuss different research paradigms and to identify the research paradigm that is suitable for this study. The research paradigms discussed included the positivistic, interpretivistic and critical research paradigms. For each of these paradigms the ontology, epistemology, methodology and characteristics of the research paradigm have been discussed. The research paradigm suitable for this study is the positivistic research paradigm, also referred to as the scientific method. The positivistic researcher follows

two basic assumptions. The first assumption is that the world is ordered and regular, not random, and the second assumption of positivism is that the researcher can investigate the world with its regular laws and patterns in an objective manner (Oates, 2006:284). The aim of positivism is to explain and measure a certain phenomenon. The characteristics of the positivistic research paradigm were similar to the research done in this study and therefore the researcher implemented a positivistic research approach.

The advantages stipulated serve as support to justify the choice of research method, data-generation method and data-analysis method. Surveys were used as research method, and questionnaires were used to generate data. Twitter data was also gathered for this study. Quantitative data was generated by means of questionnaires and was analyzed by using statistical techniques and mathematical modelling. After quantitative data analysis has been done, the researcher should interpret the results indicating what the results show and imply, how the results relate to the literature studied and the research question, and what is important in the results. The data gathered by means of questionnaires are analyzed and the researcher will attempt to answer the research question: Which techniques can be used by a company to ensure that social media is utilized effectively? The results derived from data gathered through questionnaires are discussed subsequently in Chapter 5.

CHAPTER 5: SURVEY RESULTS

5.1   Introduction

The purpose of this chapter is to present the analysis of the results drawn from the data gathered by means of questionnaires. Descriptive statistics, factor analysis, *t*-tests and ANOVA's were used, because the researcher wanted to describe some of the characteristics of the variables in the data set. In section 5.2, the descriptive statistics, factor analysis, *t*-tests and ANOVA's used in this study are discussed in greater detail.

In section 5.3, the results of section A of the questionnaire are presented. This section of the questionnaire focused on general information. In section 5.4, the results from section B of the questionnaire, which focused on different social media platforms, are presented. The results of section C of the questionnaire, which focused on social media strategies of a company, are presented in section 5.5. The results of the last section of the questionnaire, section D, are presented in section 5.6. This section focused on techniques used by companies for social media metrics. Figure 5.1 is an overview of the contents of this chapter.

Utilizing social media to the benefit of companies

Chapter 1: Problem statement

Chapter 2 - Literature Study: Social media	Chapter 3 - Literature study: Mining social media data

Chapter 4: Research design

Chapter 5: Results of questionnaire	Chapter 6: Results of Twitter data
5.1 Introduction **5.2 Descriptive statistics used in this study** **I**n §5.2.1 the descriptive statistics used in this study are describe. In §5.2.2 factor analysis is described, while in §5.2.3 *t*-tests are described. The use of ANOVA's are described in §5.2.4. Reliability and validity are discussed in §5.2.5. **5.3 Section A – General information** Results of analyzed data regarding general information of the company is given in this section. **5.4 Section B – Social media platforms** Results of analyzed data regarding social media platforms are presented in this section. **5.5 Section C – Social media strategy** Results of analyzed data regarding companies' social media strategies are presented in this section. **5.6 Section D – Social media metrics** Results of analyzed data regarding the social media metrics used by companies' are presented. **5.7 Conclusion**	

Chapter 7: Discussion, interpretation and conclusion

Figure 5.1 - Overview of Chapter 5

5.2 Statistical techniques used in this study

In this section the statistical techniques, such as descriptive statistics, exploratory factor analysis, *t*-tests and ANOVA's are discussed.

5.2.1 Descriptive statistics used in this study

Descriptive statistics focus on individual variables and characterize the distribution for each variable in the data set that the researcher wants to examine (Meyers et al., 2009:77). Meyers et al. (2009:77) divided descriptive statistics into two categories: measures of central tendency and dispersion (variability). The following descriptive statistics were used during the analyses of data gathered by means of questionnaires.

Measures of central tendency:

- Frequency: The number of occurrences in the random sample that associated with a specified category, scale or item.

- Cumulative frequency: The sum of a frequency and all the frequencies in a frequency distribution.

- Percentage: The division of the frequency by the total number of observations (respondents) and then multiplying it by 100.

- Cumulative percentage: The division of the cumulative frequency by the total number of observations (respondents) and then multiplying it by 100.

- Mean: This is the arithmetic average. The sum of scores divided by the number of occurrences with valid data entries for a variable, also known as the population (N) (Meyers et al., 2009:77). Mean (μ) can be calculated by using the following equation:

$$\overline{x} = \frac{\sum x_i}{n}$$

Measures of dispersion:

- Minimum and maximum: The lowest and highest values in the distribution (Meyers et al., 2009:78). The minimum and maximum for closed questions ranged from 1 to 4.

- Standard deviation: The standard deviation is a measure of the dispersion of the data around a mean or the average value (Taylor & Cihon, 2004:23). Standard deviation (σ) can be calculated by using the following equation:

$$s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n - 1}}$$

Other statistical techniques used include exploratory factor analysis, *t*-Tests and ANOVA's. These terms will be discussed subsequently.

5.2.2 Exploratory factor analysis

Factor analysis is described as a variable-reduction procedure. Relations among variables can be summarized when a number of variables are replaced with a few factors (Goldberg & Velicer, 2006:209; Ho, 2006:203). Goldberg and Velicer (2006:209) describe exploratory factor analysis as a procedure for identifying summary constructs when the nature of these constructs is still unknown.

There are three basic steps to factor analysis (Ho, 2006:204):

1. Calculating the correlation matrix of all variables
2. Extraction of initial factors (deciding on the number of factors)
3. Rotation of the extracted factors

When deciding on the number of factors, eigenvalues are taken into consideration and are produced by means of principal components analysis. Eigenvalues are a representation of the variance of each factor. All factors with eigenvalues of 1 or greater will be retained (Ho, 2006:207). Another approach used to determine the number of factors is known as the Kaiser-Guttman measure of sampling adequacy. After the researcher has decided on the number of factors, the loadings of each factor needed to be determined and for this study Cronbach's alpha was used.

Cronbach's alpha measures how closely related a set of items as a group is; in other words, it is a measure of internal consistency. Cronbach's alpha is not only perceived as a statistical test, but also as a measure of scale reliability. A Cronbach alpha value that is less than 0.5 is considered to be poor and unacceptable (Gliem & Gliem, 2003:87). The correlation coefficients (factor loadings) between the variables and the factors will contribute to interpreting the factors. Correlation coefficients greater than ± 0.33 are considered to meet the minimum level of practical significance (Ho, 2006:207; Roberts & Priest, 2006:42).

5.2.3 *t*-Tests

The two-sample *t*-Test procedure is used to determine if the means of two independent distributions are considerably different (Ho, 2006:41; Meyers *et al.*, 2009:195). When two sample *t*-tests are done, the following assumptions are made (Ho, 2006:41):

- The two samples are independent from each other.
- Both samples have a normal distribution.
- One of the variables is continuous dependent and the other is categorical independent.

5.2.4 ANOVA's

Meyers *et al.* (2009:213) explain that an ANOVA (analysis of variance) can be used to determine if the means of two or more distributions are significantly different. The test statistic that is calculated when using the ANOVA procedure is an *F* ratio (Meyers *et al.*, 2009:215). It is assumed that the sampling distributions of means are normal within each group (Tabachnick & Fidell, 2013:204). If the *p*-value is below 0.05 there is a statistically significant difference between the group means (the variances are unequal). Tukey's studentized range test indicates the specific groups that differed. A limitation of ANOVA's is that the effect size (the size of the difference) is not always indicated.

For this study, Cohen's *d* was calculated to examine the effect size (Meyers *et al.,* 2009:200). Cohen's *d* is calculated as follows: $d = \frac{absolute\ mean\ difference}{average\ standard\ deviation}$
Cohen proposed that the value of *d* = 0.2 can be thought of as a small effect size, while *d* = 0.5 is a medium effect size and 0.8 a large effect size (Meyers *et al.,* 2009:200). Cohen's *d*-value is also used in *t*-Tests.

5.2.5 Reliability and validity

By performing specific statistical tests, researchers can measure the reliability and validity of research (Roberts & Priest, 2006:42). Reliability is used to describe how a data collection method, for example a questionnaire, can produce similar results within different circumstances (Roberts & Priest, 2006:41). Heale and Twycross (2015:66) are of the opinion that reliability relates to the consistency of a measure. Individual questions in a questionnaire can be measured by using statistical tests, such as Cronbach's alpha.

Heale and Twycross (2015:67) describe the following attributes of reliability: internal consistency, stability and equivalence. Reliability is often associated with internal consistency, which is the relationship between all the results obtained from a single test or a survey (Roberts & Priest, 2006:42). Internal consistency is also described as the extent to which items on a scale measure one construct (Heale & Twycross 2015:67). Stability refers to whether results can be consistent when tested continuously. Equivalence is described as the consistency between the responses of multiple respondents (Heale & Twycross, 2015:67).

There are two broad measures of validity, namely external validity and internal validity (Roberts & Priest, 2006:43). These two terms were also discussed in §4.2.5. Heale and Twycross (2015:66) describe three types of validity: content validity, construct validity and criterion validity. Content validity refers to the extent to which a research method accurately measures all aspects of a construct. Construct validity is the extent to which a research method, for example a survey, measures the intended construct. Criterion validity is concerned with how

related the research method is to other methods that measure similar variables (Heale & Twycross, 2015:66).

Good quality research provides evidence of how reliability and validity have been tested. This evidence also influences the decision on whether the findings should be applied (Heale & Twycross, 2015:67).

The results obtained from data gathered by means of the questionnaire are presented next. The results presented in this chapter are discussed in Chapter 7.

5.3 Section A – General information

The aim of this section is to describe and present the statistical results of general information gathered regarding companies. The results obtained in this section will also be used to describe the population. Table 5.1 presents the type of data gathered for each question, the statistical techniques used to analyse the data and the size of the population (respondents who answered the question). The size of the population will be indicated with an N.

Table 5.1 - Section A: Statistical techniques used

Question	Type of data gathered	Statistical techniques used	N	Depicted in
1. In which of the following business areas are you mainly employed in?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	122	Table 5.2.
2. Roughly how many full-time employees are currently working for the company?	Nominal data		122	Table 5.3.
3. How useful do you think that social media has been for the company?	Ordinal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	122	- Table 5.4. - Figure 5.2. - Table 5.5.

5.3.1 Business area

The respondents were asked to indicate the business area in which they are mainly employed. The most respondents indicated that they are mainly employed in information technology and communications (53.28%), while the minority of respondents are employed in a legal business area (4.10%). Table 5.2 presents the frequency and percentage of respondents respectively.

Table 5.2 - Results obtained regarding business area

Business are	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Retail, sales and marketing	18	18	14.75	14.75
2. Banking, finance and accounting	20	38	16.39	31.15
3. Insurance	8	46	6.56	37.70
4. Information technology and communications	65	111	53.28	90.98
5. Legal	5	116	4.10	95.08
6. Government	6	122	4.92	100.00

5.3.2 Size of company

The respondents were asked to indicate the size of the company at which they are currently employed. The size of a company is demonstrated by the number of people employed, which can then be categorized into a specific enterprise. The dominant size is 10 – 49 employees (small enterprise), while the company size least indicated by respondents is 1 – 9 employees (micro-enterprise). The medium-sized enterprise (50- 249 employees) and the large enterprise (250+ employees) almost had similar frequencies in this study. The results obtained regarding company size are presented in Table 5.3.

Table 5.3 - Results obtained regarding company size

Company size	Frequency:	Cumulative frequency	Percentage	Cumulative percentage
1. 1 – 9 employees (micro-enterprise)	15	15	12.30	12.30
2. 10 – 49 employees (small enterprise)	46	61	37.70	50.00
3. 50 – 249 employees (medium-sized enterprise)	31	92	25.41	75.41
4. 250+ employees (large enterprise)	30	122	24.59	100.00

5.3.3 Usefulness of social media for the company

The researcher wanted to gather the opinion of each respondent regarding the usefulness of social media for the company for which they work. Respondents were asked to express their opinion on a Likert scale question ranging from not useful to very useful. Most of the respondents (45.08%) were of the opinion that social media has been useful for the company. Only 31.15% of the respondents were of the opinion that social media has not been useful or moderately useful for the company. The results regarding the usefulness of social media for the company are presented in Figure 5.2 and Table 5.4. The mean and standard deviation is presented in Table 5.5.

Table 5.4 - Results regarding the usefulness of social media for the company

Usefulness of social media for the company	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Not useful	12	12	9.84	9.84
2. Moderately useful	26	38	21.31	31.15
3. Useful	55	93	45.08	76.23
4. Very useful	29	122	23.77	100.00
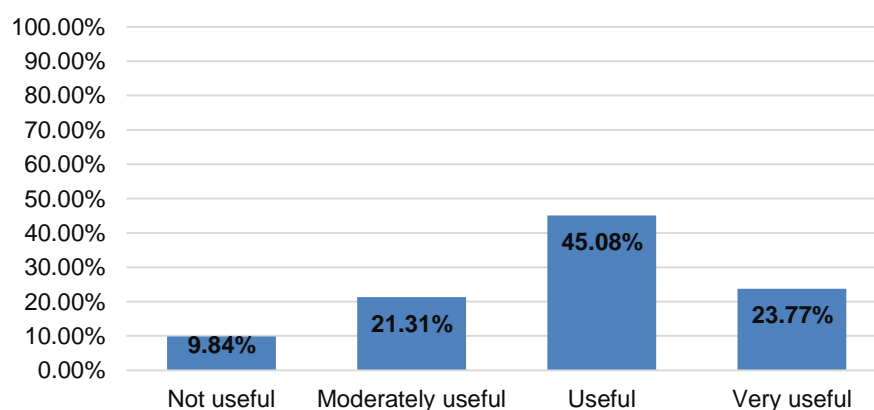
Figure 5.2 - Opinions of respondents regarding the usefulness of social media for companies

Table 5.5 - Mean and standard deviation of the usefulness of social media for companies

Mean	Standard deviation
2.83	0.91

This section presented general information regarding a company and the employees who acted as respondents/participants in this study. The subsequent section will describe the results obtained from Section B of the questionnaire regarding social media platforms.

5.4   Section B – Social media platforms

The aim of this section is to describe and present the statistical results of the data gathered regarding the use of social media platforms in companies. Table 5.6 presents the type of data gathered for each question, the statistical techniques used to analyse the data and the size of the population (respondents who answered the question).

Table 5.6 - Section B: Statistical techniques used

Question	Type of data gathered	Statistical techniques used	N	Depicted in
1. Are you allowed to visit external social media platforms, e.g. Facebook, Twitter, LinkedIn, etc. during working hours?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	122	Table 5.7.
2. Does the company have social media training across the company to ensure that employees understand how to engage on social media platforms and consistently represent the company?	Nominal data		122	Table 5.8.
3. How long has the company been using social media platforms?	Nominal data		122	- Table 5.9. - Figure 5.3.
4.1. Please rate the effectiveness of the following social media platforms used for marketing by the company?	Ordinal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	Population size is given for each item in Table 5.10, Table 5.11. and Table 5.12.	- Table 5.10. - Figure 5.4.
4.2. Please rate the effectiveness of the following social media platforms used for branding by the company?				- Table 5.11. - Figure 5.5.
4.3. Please rate the effectiveness of the following social media platforms used for customer relationship management by the company?				- Table 5.12. - Figure 5.5.
4.4. Are there any other social media platforms used by the company that is not listed in the previous questions?	Text	• Textual data was not analyzed by using statistical techniques, but by means of theme analysis and coding.	13	Discussed in §5.4.4.4
5.1. How does the company determine awareness of customers and clients regarding social media platforms being used?	Text	• Theme analysis/coding	42	Discussed in §5.4.5.1
5.2. According to the company's perspective, rate how aware customers and clients are regarding social media platforms being used.	Ordinal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	122	Table 5.13.
6.1. According to your opinion rate each of the following uses of social media by the company.	Ordinal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	Population size is given for each item in Table 5.14.	Table 5.14.
6.2. List any other uses of social media platforms which is not listed above.	Text	• Theme analysis/coding	14	Discussed in §5.4.6.2
7. Indicate your level of agreement regarding the following statements.	Ordinal data	• Frequency • Cumulative frequency	Population size is	Table 5.15 to table 5.23.

Question	Type of data gathered	Statistical techniques used	N	Depicted in
		• Percentage • Cumulative percentage • Mean • Standard deviation	given for each item in Table 5.15.	
8. Indicate your level of agreement with the following statement: This social media platform allows a customer to express sentiment and his/her opinion.	Ordinal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	Population size is given for each item in Table 5.24.	Table 5.24
9. Is your company paying to promote posts on social media platforms?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	122	Table 5.25
10.1. According to your opinion does different social media platforms attract a different audience?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	122	Table 5.26
10.2. How does the company determine the audience of their social media platforms?	Text	• Theme analysis/coding	41	Discussed in §5.4.10.2
11. What is a limitation that the company has experienced on a social media platform and how did the company overcome this?	Text	• Theme analysis/coding	28	Discussed in §5.4.11
12.1. Has the company hired or appointed employees specifically to manage the company's social media platforms?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	122	Table 5.27
12.2. If you have answered Yes to the question 12.1. have the employees appointed to manage the social media platforms received training or own a certain qualification, if Yes please specify.	Text	• Theme analysis/coding	27	Discussed in §5.4.12.2
12.3. Are these employee(s) appointed in-house or outsourced?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	47	Table 5.28

Subsequently, the results of Section B questions (1 – 12.3.), given in Table 5.6, will be presented.

5.4.1 Visitation of external social media platforms

Respondents were asked whether they are allowed to visit external social media platforms, for example Facebook, Twitter, LinkedIn, etc. during working hours. The majority of respondents – 81.15% (99 respondents) are allowed to visit social media platforms during working/office hours. Table 5.7 presents the results regarding the visitation of external social media platforms during working hours.

Table 5.7 - Visitation of external social media platforms

Visitation of external social media platforms	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Yes	99	99	81.15	81.15
2. No	23	122	18.85	100.00

5.4.2 Training

Respondents were asked to indicate whether the company provides training to ensure that employees understand how to engage on social media platforms. It is important for an employee to understand that the company should be represented appropriately on social media platforms to avoid a social media crisis. The majority of respondents – 70.49% (86 respondents) indicated that no training was provided to employees regarding the presentation of the company on social media platforms.

Table 5.8 - Social media training within the company

Social media training within the company	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Yes	36	36	29.51	29.51
2. No	86	122	70.49	100.00

5.4.3 Period of time using social media

The respondents were asked to indicate the period of time that the company has been using social media. Respondents could choose from six different timeframes. Of the respondents 29.51% (36 respondents) indicated that the company has been using social media platforms for the past 1 – 2 years. 27.87% (34 respondents) indicated that the company has been using social media for the past 9 – 12 months. The time frame, 1 – 3 months, was indicated by only 1.64% of the respondents (two respondents). Only 17.21% of the respondents (21 respondents) indicated that the company has been using social media for more than five years. Table 5.9 presents the results regarding the period of time that the company has been using social media. Figure 5.3 is a representation of the results given in Table 5.9.

Table 5.9 - Period of time that the company has been using social media

Period of time that the company has been using social media	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. 1 – 3 months	2	2	1.64	1.64
2. 4 – 8 months	6	8	4.92	6.56
3. 9 – 12 months	34	42	27.87	34.43
4. 1 – 2 years	36	78	29.51	63.93
5. 3 – 4 years	23	101	18.85	82.79
6. 5+ years	21	122	17.21	100.00
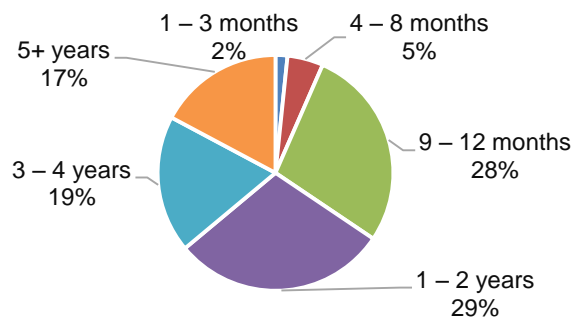
Figure 5.3 - Period of time that a company has been using social media

5.4.4 Effectiveness of social media platforms

The effectiveness of social media platforms for marketing, branding and customer relationship management will be discussed separately in relation to eight different platforms: Facebook, Twitter, LinkedIn, Google+, YouTube, Pinterest, Instagram and Tumblr.

5.4.4.1 Effectiveness of social media platform for marketing

Respondents were asked to rate each of the social media platforms mentioned above (if applicable) regarding the effectiveness of the platform when used for marketing. Table 5.10 presents the population (N), frequency, percentage, mean and standard deviation of each social media platform.

Table 5.10 - Descriptive statistics regarding the effectiveness of social media platforms for marketing

Platform	N		Frequency	Percentage	Mean	Standard deviation
Facebook	119	1. Not effective	21	17.65	**2.57**	0.93
		2. Moderately effective	25	21.01		
		3. Effective	57	47.90		
		4. Very effective	16	13.45		
LinkedIn	116	1. Not effective	24	20.69	2.56	1.04
		2. Moderately effective	27	23.28		
		3. Effective	41	35.34		
		4. Very effective	24	20.69		
Twitter	117	1. Not effective	38	32.48	2.13	0.96
		2. Moderately effective	35	29.91		
		3. Effective	35	29.91		
		4. Very effective	9	7.69		
YouTube	101	1. Not effective	40	39.60	2.08	1.08
		2. Moderately effective	28	27.72		
		3. Effective	18	17.82		
		4. Very effective	15	14.85		
Instagram	100	1. Not effective	48	48.00	2.00	1.08
		2. Moderately effective	17	17.00		
		3. Effective	22	22.00		
		4. Very effective	13	13.00		
Google+	113	1. Not effective	48	42.48	1.90	0.92
		2. Moderately effective	33	29.20		
		3. Effective	27	23.89		
		4. Very effective	5	4.42		
Pinterest	95	1. Not effective	56	58.95	1.66	0.92
		2. Moderately effective	20	21.05		
		3. Effective	14	14.74		
		4. Very effective	5	5.26		
Tumblr	92	1. Not effective	60	65.22	1.57	0.91
		2. Moderately effective	18	19.57		
		3. Effective	8	8.70		
		4. Very effective	6	6.52		

Figure 5.4 presents the results given in Table 5.10 according to the percentage of the population who indicated the level of effectiveness. The perception of respondents was that

Facebook is effective when used for marketing with a mean of 2.57, while Tumblr is moderately effective with a mean of 1.57.
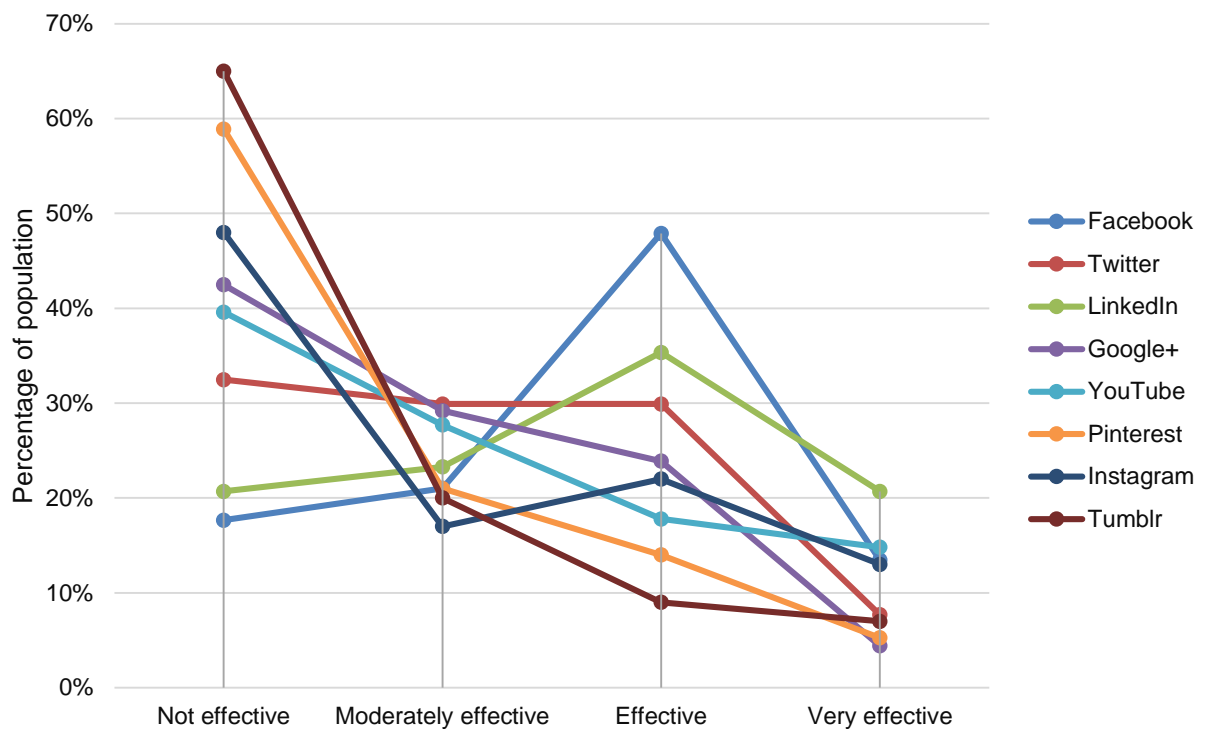
Figure 5.4 - Effectiveness of social media when used for marketing in relation to different social media platforms

The results regarding the effectiveness of social media platforms for branding will be presented next.

5.4.4.2   Effectiveness of social media platform for branding

Respondents were asked to rate each of the social media platforms regarding the effectiveness of the platform when used for branding. Table 5.11 presents the descriptive statistics regarding the effectiveness of social media platforms when used for branding.

Table 5.11 - Descriptive statistics regarding the effectiveness of social media platforms for branding

Platform	N		Frequency	Percentage	Mean	Standard deviation
Facebook	107	1. Not effective	30	28.04	**2.49**	1.10
		2. Moderately effective	16	14.95		
		3. Effective	40	37.38		
		4. Very effective	21	19.63		
Twitter	103	1. Not effective	26	25.24	2.43	1.04
		2. Moderately effective	24	23.30		
		3. Effective	36	34.95		
		4. Very effective	17	16.50		
LinkedIn	99	1. Not effective	31	31.31	2.32	1.09
		2. Moderately effective	21	21.21		
		3. Effective	31	31.31		
		4. Very effective	16	16.16		

Platform	N			Frequency	Percentage	Mean	Standard deviation
Pinterest	96		1. Not effective	44	45.83	2.04	1.15
			2. Moderately effective	21	21.88		
			3. Effective	14	14.58		
			4. Very effective	17	17.71		
Instagram	97		1. Not effective	42	43.30	1.98	1.01
			2. Moderately effective	23	23.71		
			3. Effective	24	24.74		
			4. Very effective	8	8.25		
Google+	102		1. Not effective	49	48.04	1.81	0.90
			2. Moderately effective	26	25.49		
			3. Effective	24	23.53		
			4. Very effective	3	2.94		
Tumblr	94		1. Not effective	52	55.32	1.81	1.06
			2. Moderately effective	19	20.21		
			3. Effective	12	12.77		
			4. Very effective	11	11.70		
YouTube	102		1. Not effective	47	46.08	**1.78**	0.90
			2. Moderately effective	37	36.27		
			3. Effective	11	10.78		
			4. Very effective	7	6.86		

Figure 5.5 presents the results given in Table 5.11 according to the percentage of the population who indicated the level of effectiveness. The perception of respondents was that Facebook is effective (mean of 2.49) when used for branding while YouTube is moderately effective (mean of 1.78).
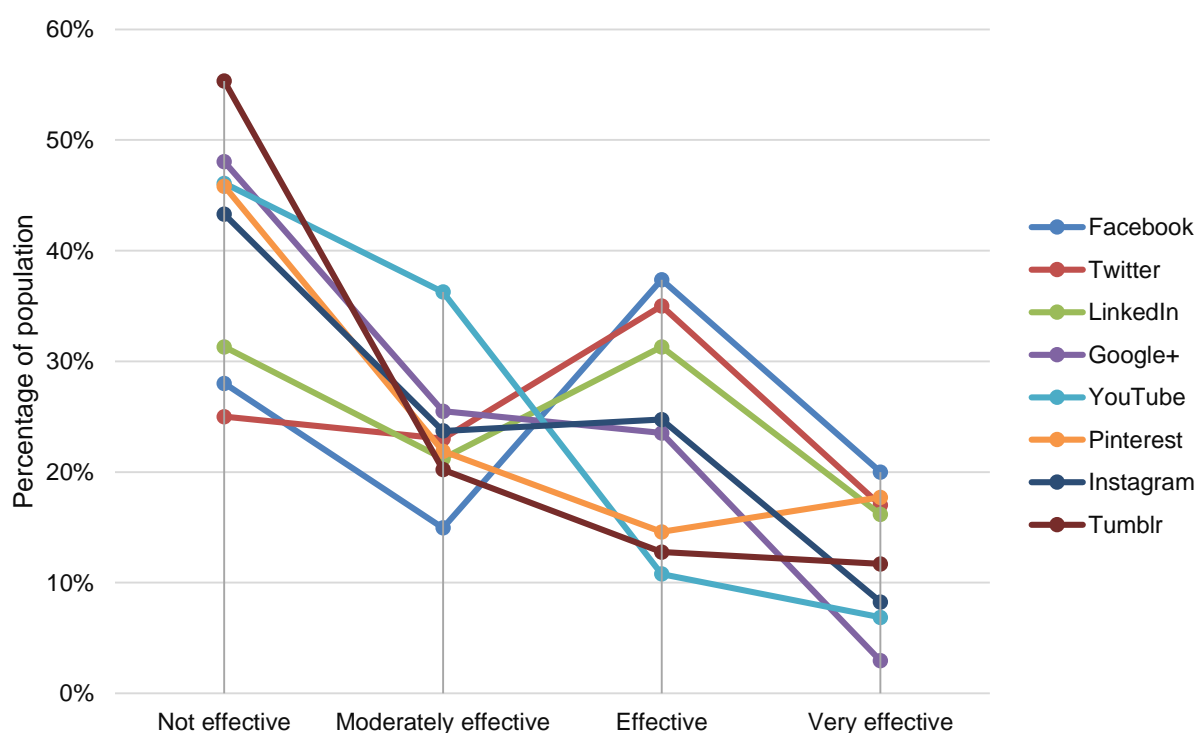
Figure 5.5 - Effectiveness of social media when used for branding in relation to different social media platforms

The results regarding the effectiveness of social media platforms for customer relationship management are presented next.

5.4.4.3 Effectiveness of social media platform for customer relationship management

Respondents were asked to rate each of the social media platforms regarding the effectiveness of the platform when used for customer relationship management. In Table 5.12, the population (N), frequency, percentage, mean and standard deviation of each social media platform are presented. In Table 5.12 it can be seen that the social media platform that is considered to be most effective for customer relationship management is Facebook (mean of 2.38), followed by Twitter (mean of 2.26). The social media platform considered not effective for customer relationship management is Instagram with a mean of 1.84.

Table 5.12 - Descriptive statistics regarding the effectiveness of social media platforms for customer relationship management

Platform	N		Frequency	Percentage	Mean	Standard deviation
Facebook	100	1. Not effective	31	31.00	**2.38**	1.12
		2. Moderately effective	19	19.00		
		3. Effective	31	31.00		
		4. Very effective	19	19.00		
Twitter	97	1. Not effective	33	34.02	2.26	1.09
		2. Moderately effective	21	21.65		
		3. Effective	28	28.87		
		4. Very effective	15	15.46		
LinkedIn	97	1. Not effective	37	38.14	2.16	1.10
		2. Moderately effective	21	21.65		
		3. Effective	25	25.77		
		4. Very effective	14	14.43		
YouTube	96	1. Not effective	41	42.71	2.01	1.07
		2. Moderately effective	26	27.08		
		3. Effective	16	16.67		
		4. Very effective	13	13.54		
Google+	92	1. Not effective	46	50.00	1.96	1.12
		2. Moderately effective	17	18.48		
		3. Effective	16	17.39		
		4. Very effective	13	14.13		
Pinterest	92	1. Not effective	47	51.09	1.92	1.11
		2. Moderately effective	18	19.57		
		3. Effective	14	15.22		
		4. Very effective	13	14.13		
Tumblr	94	1. Not effective	48	51.06	1.89	1.06
		2. Moderately effective	18	19.15		
		3. Effective	18	19.15		
		4. Very effective	10	10.64		
Instagram	87	1. Not effective	46	52.87	**1.84**	1.04
		2. Moderately effective	18	20.69		
		3. Effective	14	16.09		
		4. Very effective	9	10.34		

Figure 5.6 presents the results given in Table 5.12 according to the percentage of the population who indicated the effectiveness of a specified social media platform for customer relationship management.
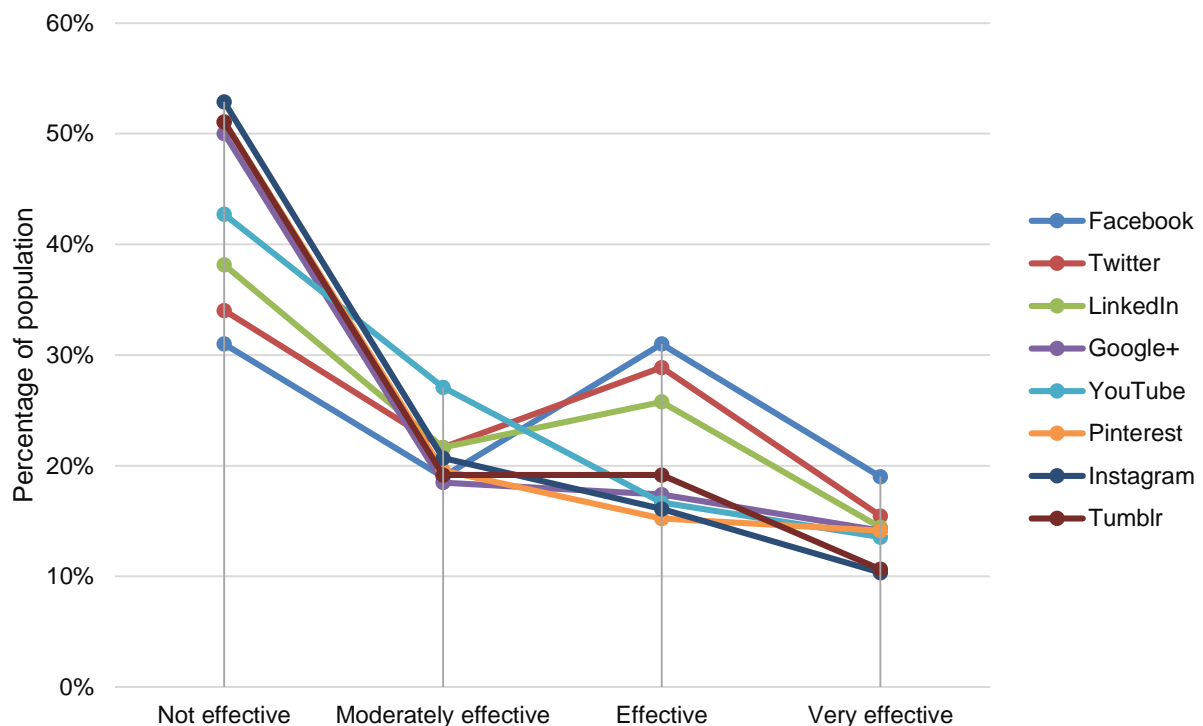
Figure 5.6 - Effectiveness of social media when used for customer relationship management in relation to different social media platforms

Respondents are of the opinion that Facebook is effective for marketing (mean of 2.57), branding (mean of 2.49) and moderately effective for customer relationship management (mean of 2.38). For all of these three purposes Facebook was ranked the highest. Respondents are of the opinion that Twitter is moderately effective for branding (mean of 2.43) and for customer relationship management (mean of 2.26). Respondents indicated that LinkedIn is more effective for marketing (mean of 2.56) than Twitter (mean of 2.13).

Respondents were asked to indicate any other social media platforms used for marketing, branding or customer relationship management.

5.4.4.4 Other social media platforms

Only 13 respondents answered this open-ended question. Most of the respondents mentioned Skype. People might not always consider Skype to be a social media platform, but it does fulfil some of the same functions that other social media platforms offer. Another social media network mentioned was Yammer, which is used in collaboration with Office 365.

Respondents also mentioned WhatsApp. WhatsApp is a messaging platform, but can be used for marketing, branding or customer relationship management. Respondents also mentioned that bulk SMS campaigns are used and that the company web site is linked to social media platforms where the company has a presence. The results regarding the awareness of customers will subsequently be discussed.

5.4.5 Awareness of customers

The following results indicate how companies, within the population, determine the awareness of their customers/clients regarding the company's use of social media platforms. Respondents also indicated the level of awareness of customers regarding the company's social media platforms. First, the results obtained regarding the techniques used to determine the awareness of customers/clients will be presented.

5.4.5.1   Determination of awareness

The respondents were asked to indicate how the company determines the awareness of customers and clients regarding social media platforms being used. Only 42 respondents answered this open-ended question. Most of the respondents indicated that the company makes use of Google analytics, Facebook reports and Hootsuite. Respondents also indicated that the company makes use of web site and social media platform tracking. Another measurement of awareness that companies use is the number of likes on their Facebook and YouTube channels.

The following are two of the responses that carried meaning to the question:

Through the number of inquiries on the specific social media platform and the general increase of inquiries at head office where clients and/or customer specify they have seen the company on a social media platform.

Determine, or create? We embed code into web sites and use tracking cookies to feed ads to individuals that use our services. We can then track the usage back into our database for further analysis.

Respondents also mentioned that the company makes use of questionnaires and surveys to determine how aware their clients are regarding the company's presence on social media platforms.

5.4.5.2   Company's perspective of awareness of customers

Respondents were asked to rate how aware customers and clients are regarding social media platforms being used. The calculated mean of 2.57 tends to 3, indicating that customers/clients are aware of social media platforms being used by the company. Only 11.48% of the

respondents (14 respondents) are of the opinion that the company's customers/clients are not aware of the social media platforms used. 39.34% of the respondents (48 respondents) are of the opinion that the company's customers/clients are aware of the social media platforms used by the company. Only 14.75% of the respondents (18 respondents) are of the opinion that customers/clients level of awareness is extreme. Table 5.13 indicates the descriptive statistics of the awareness of customers/clients regarding social media platforms being used.

Table 5.13 – Descriptive statistics of the awareness of customers/clients regarding social media platforms being used

Company's perspective of awareness of customers	Frequency	Cumulative frequency	Percentage	Cumulative percentage	Mean	Standard deviation
1. Not aware	14	14	11.48	11.48		
2. Moderately aware	42	56	34.43	45.90	2.57	0.88
3. Aware	48	104	39.34	85.25		
4. Extremely aware	18	122	14.75	100.00		

Different uses of social media were identified during the literature review which led to the following questions on the uses of social media.

5.4.6   Uses of social media

The aim of the following two questions was to identify different uses of social media within companies. Firstly, seven different reasons, were identified from literature. Respondents were also asked to indicate other uses of social media not listed.

5.4.6.1   Rating regarding the uses of social media

Respondents were asked to rate each of the listed uses of social media by the company on a Likert scale ranging from 1 – 4. In Table 5.14 the descriptive statics regarding different uses of social media platforms in companies are presented. The most popular use (mean of 2.62) was to increase brand awareness. To offer promotional offers/items was the least popular use of social media platforms (mean of 2.10).

26.23% of the respondents (32 respondents) indicated that social media platforms are almost always used to increase brand awareness. 33.61% of the respondents (41 respondents) indicated that social media platforms are regularly used to advertise products or services. For this study it is also important to investigate the feedback of customers and customer relationship management. Only 19.67% of the respondents (24 respondents) indicated that social media platforms are almost always used to gain feedback from customers. 42.62% of the respondents (52 respondents) indicated that social media platforms are regularly and almost always used to engage with customers.

Table 5.14 - Descriptive statistics regarding different uses of social media platforms in companies

Use	N		Frequency	Percentage	Mean	Standard deviation
To increase brand awareness	122	1. Almost never used	27	22.13	**2.62**	1.10
		2. Moderately used	24	19.67		
		3. Regularly used	39	31.97		
		4. Almost always used	32	26.23		
To advertise products/services	122	1. Almost never used	23	18.85	2.56	1.02
		2. Moderately used	33	27.05		
		3. Regularly used	41	33.61		
		4. Almost always used	25	20.49		
For business to business purposes e.g. LinkedIn	122	1. Almost never used	28	22.95	2.48	1.05
		2. Moderately used	31	25.41		
		3. Regularly used	40	32.79		
		4. Almost always used	23	18.85		
To gain feedback from customers	122	1. Almost never used	34	27.87	2.38	1.09
		2. Moderately used	32	26.23		
		3. Regularly used	32	26.23		
		4. Almost always used	24	19.67		
To engage (in conversation) with customers	121	1. Almost never used	32	26.23	2.34	1.05
		2. Moderately used	38	31.15		
		3. Regularly used	31	25.41		
		4. Almost always used	21	17.21		
To analyze the competition in the industry	122	1. Almost never used	37	30.58	2.31	1.09
		2. Moderately used	32	26.45		
		3. Regularly used	30	24.79		
		4. Almost always used	22	18.18		
To offer promotional offers/items	122	1. Almost never used	44	36.07	**2.10**	1.04
		2. Moderately used	38	31.15		
		3. Regularly used	24	19.67		
		4. Almost always used	16	13.11		

Respondents were asked to name and describe other uses of social media platforms not listed in Table 5.14 which led to the following results.

5.4.6.2 Other uses of social media

Only 14 respondents answered this open-ended question. The majority of the respondents mentioned the use of social media platforms for community engagement, talent recruitment and distributing information regarding the company. One of the respondents mentioned that the company uses social media platforms to gather information about their customers/clients for internal analysis.

Next, the results regarding the effectiveness of social media platforms for the different uses listed in Table 5.14 will be presented.

5.4.7 Effectiveness of social media

Respondents were asked to indicate their level of agreement regarding a list of statements as given in Table 5.15. In Table 5.15, the descriptive statistics regarding the effectiveness of social media platforms for a specified purpose are presented.

Table 5.15 - Descriptive statistics regarding the effectiveness of social media platforms for a specified purpose

Use	Variable	N		Frequency	Percentage	Mean	Standard deviation
Social media is effective for advertising a product/service.	B_7_4	121	1. Strongly disagree	12	9.92	2.94	0.95
			2. Disagree	22	18.18		
			3. Agree	48	39.67		
			4. Strongly agree	39	32.23		
Social media has revolutionized the marketing of a product/service.	B_7_2	122	1. Strongly disagree	16	13.11	2.93	1.00
			2. Disagree	16	13.11		
			3. Agree	51	41.80		
			4. Strongly agree	39	31.97		
Social media is effective for selling a product/service.	B_7_1	122	1. Strongly disagree	15	12.30	2.84	0.96
			2. Disagree	22	18.03		
			3. Agree	53	43.44		
			4. Strongly agree	32	26.23		
Social media is effective for brand awareness.	B_7_5	122	1. Strongly disagree	18	14.75	2.83	0.99
			2. Disagree	17	13.93		
			3. Agree	54	44.26		
			4. Strongly agree	33	27.05		
Social media is great for engaging with customers.	B_7_3	122	1. Strongly disagree	12	9.84	2.80	0.91
			2. Disagree	28	22.95		
			3. Agree	54	44.26		
			4. Strongly agree	28	22.95		

Table 5.16 – Eigenvalues of variables for variables B_7_1 to B_7_5

	Variable	Eigenvalue	Proportion
	B_7_4	1.902	38.04%
	B_7_2	0.928	18.55%
Effectiveness	B_7_1	0.855	17.10%
	B_7_5	0.735	14.69%
	B_7_3	0.580	11.61%

In Table 5.16, the variable, eigenvalue of the variable, as well as proportion are given. The 5-item scale has an eigenvalue that indicates the amount of variation in the items accounted for by each variable. For example, the first variable has an eigenvalue of 1.902, meaning it accounts for 38.04% of the variance (1.902/5 = 0.3804). A factor with an eigenvalue ≥ 1 explains more variance than a single observed variable. This indicated 1 factor, which was identified as effective. The Cronbach alpha obtained was 0.59, indicating acceptable validity and reliability of the factor, as can be viewed in Table 5.17 (Gliem & Gliem, 2003:87).

Table 5.17 - Cronbach alpha obtained for effectiveness factor

Factor:	N	Mean	Standard deviation	Cronbach alpha
Effectiveness	122	2.87	0.59	0.59

Table 5.18 indicates the MSA (Measurement of Sampling Adequacy) and communality for each of the variables in Table 5.15 and Table 5.16. The communality estimates indicate the percentage of variance explained by the identified factor. The overall MSA of 0.673 also indicates good sampling adequacy. The final communality estimate total was 1.90 with the one factor.

Table 5.18 - MSA and communality for variables B_7_1 to B_7_5

Variable	MSA	Communality
B_7_1	0.731	0.402
B_7_2	0.658	0.455
B_7_3	0.659	0.250
B_7_4	0.649	0.491
B_7_5	0.683	0.304

A two-sample *t*-test was done using the effectiveness factor. Two groups were identified. Group 1 indicates the banking, finance and accounting business area, while Group 2 indicates the information technology and communications business area. These two groups were chosen, because the majority of data gathered by means of questionnaires fell within these two business areas. The results of the *t*-Test are presented in Table 5.19.

Table 5.19 - *t*-Test done using effectiveness factor on business areas

Factor	Group	N	Mean	Standard deviation	p-value (when random sampling is assumed)	d-value
Effectiveness	1	20	3.04	0.64	0.15	-0.38
	2	65	2.80	0.60		

Note: Group 1 – Banking, finance and accounting.
         Group 2 – Information technology and communications.
* Statistically significant at 0.05 level according to t-test results for independent groups
** Statistically significant at 0.01 level according to t-test results for independent groups

$^\Delta$ Medium effect in practice
$^{\Delta\Delta}$ Large effect in practice and practically significant

From Table 5.19 the results indicate that respondents whom are employed in the banking, finance and accounting business area are of the opinion that social media is effective, with a mean of 3.04, when used for advertising a product or service. Respondents employed in the information technology and communications business area are also of the opinion that social media is effective, mean of 2.80, for advertising a product or service.

Correlations were made between different aspects of the questionnaire and the two identified groups. Table 5.20 presents the effect sizes regarding the two business areas measured according to different aspects. From the results given in Table 5.20, both of the business areas (banking, finance and accounting, as well as information technology and communications) used for the *t*-test are of the opinion that social media is useful for the company.

From the results presented in Table 5.20, respondents whom are employed in the banking, finance and insurance business area are of the opinion that their customers or clients are

aware of the company's social media presence (mean of 2.80). Respondents employed in the information technology and communications business area are of the opinion that their customers or clients are only moderately aware of the company's social media presence (mean of 2.42).

The respondents employed in the banking, finance and accounting business area are of the opinion that the company's social media goals are integrated with the overall business objectives (mean of 2.88). Respondents whom are employed in the information technology and communications business area are of the opinion that the company's social media goals are only moderately integrated with the company's business objectives (mean of 2.47). These results are also presented in Table 5.20.

From the results presented in Table 5.20, respondents employed in the banking, finance and accounting business area indicated that the social media posts are made less often. Respondents who are employed in the information technology and communications indicated that social media posts are made more often. Respondents employed in the banking, finance and accounting business area indicated that Twitter posts should be linked to Facebook (mean of 2.55), while respondents employed in the information technology and communications business area are of the opinion that linking Twitter posts to Facebook should be limited to an extent.

Table 5.20 – Effect sizes regarding two business areas for effectiveness

Item	*Group*	**N**	*Mean*	*Standard deviation*	**p-value** (when random sampling is assumed)	**d-value**
Usefulness of social media for company.	1	20	3.05	0.76	0.32	-0.23
	2	65	2.85	0.91		
Awareness of customers/clients	1	20	2.80	0.83	0.08	-0.45
	2	65	2.42	0.85		
Integration of social media goals with overall business objectives.	1	8	2.88	0.83	0.26	-0.41
	2	38	2.47	0.98		
Number of times content is posted on social media platforms.	1	20	2.55	1.28	0.17	0.35
	2	65	3.00	1.21		
Linking Twitter posts to Facebook	1	20	2.55	1.10	0.19	-0.33
	2	65	2.19	0.95		

Note: Group 1 – Banking, finance and accounting.
     Group 2 – Information technology and communications.
* Statistically significant at 0.05 level according to t-test results for independent groups
** Statistically significant at 0.01 level according to t-test results for independent groups

△ Medium effect in practice
△ △ Large effect in practice and practically significant

Another two-sample *t*-test was done using the effectiveness factor. Two groups were identified. Group 1 indicates companies that have been using social media platforms for nine months to two years, while Group 2 indicates companies that have been using social media for three to five+ years. These two groups were categorized as follows, because the majority of data

gathered by means of questionnaires fell within these two time frames. The results of the *t*-Test are presented in Table 5.21.

From the results presented in Table 5.21 respondents who indicated that the company has been using social media platforms for a time of 9 months to 2 years found that social media platforms is effective when used for advertising. The respondents whom indicated that the company has been using social media for a time of 3 to 5+ years found that social media is effective when used for advertising a product or service.

Table 5.21 - *t*-Test done using effectiveness factor on timeframe that the company has been using social media platforms

Factor	*Group*	N	*Mean*	*Standard deviation*	p-value (when random sampling is assumed)	d-value
Effectiveness	1	70	2.75	0.53	0.08	0.46
	2	44	3.06	0.67		

Note: Group 1 – 9 months to 2 years.
      Group 2 – 3 to 5+ years.
* Statistically significant at 0.05 level according to t-test results for independent groups
** Statistically significant at 0.01 level according to t-test results for independent groups

Δ Medium effect in practice
Δ Δ Large effect in practice and practically significant

Correlations were made between different aspects of the questionnaire and the two groups identified according to a timeframe (in Table 5.21). Table 5.22 presents the effect sizes regarding the two timeframes measured according to the different aspects.

From the results presented in Table 5.22 both of the groups are of the opinion that social media is useful for the company. Respondents who indicated that the company has been using social media for a period of 3 to 5+ years, are of the opinion that their customers or clients are aware of the company's social media presence (mean of 2.64), this was also the opinion of respondents who indicated that the company has been using social media for a period of 9 months to 2 years (mean of 2.53).

Respondents who indicated that the company has been using social media for a period of 3 to 5+ years also indicated that the company's social media goals are integrated with the company's overall business objectives (mean of 2.70). Respondents who indicated that the company has been using social media for 9 months to 2 years indicated that the company's social media goals are integrated moderately with the company's business objectives. This results are also presented in Table 5.22.

Both of the identified groups indicated that the companies are posting social media content on a weekly basis. Both of the identified groups also indicated that they disagree with the statement that linking Twitter posts to Facebook is a good idea. These results are also presented in Table 5.22.

Table 5.22 – Effect sizes regarding time using social media measured against effectiveness factor

Item	Group	N	Mean	Standard deviation	p-value (when random sampling is assumed)	d-value
Usefulness of social media for company.	1	70	2.69	0.84	0.04	0.37
	2	44	3.05	0.96		
Awareness of customers/clients	1	70	2.53	0.91	0.52	0.12
	2	44	2.64	0.81		
Integration of social media goals with overall business objectives.	1	22	2.36	1.05	0.18	0.32
	2	37	2.70	0.88		
Number of times content is posted on social media platforms.	1	70	2.86	1.21	0.43	0.15
	2	44	3.05	1.28		
Linking Twitter posts to Facebook	1	70	2.33	0.93	0.50	0.12
	2	44	2.45	1.07		

Note: Group 1 – 9 months to 2 years
    Group 2 – 3 to 5+ years.
* Statistically significant at 0.05 level according to t-test results for independent groups
** Statistically significant at 0.01 level according to t-test results for independent groups

$^\Delta$ Medium effect in practice
$^{\Delta\Delta}$ Large effect in practice and practically significant

An ANOVA was done and the effect sizes regarding the number of employees were tested against the factor identified. The results of this ANOVA are presented in Table 5.23. Four groups were identified according to the type of enterprise and number of employees. Group 1 indicated as 1 – 9 employees (micro-enterprise), Group 2 was indicated as 10 – 49 employees (small enterprise), Group 3 indicated 50 – 249 employees (medium enterprise) while Group 4 indicated 250+ employees (large enterprise). There were no significant differences between these four groups and the effectiveness factor identified.

Table 5. 23 – Effect sizes regarding number of employees for effectiveness

Construct	Group	N	Mean	Standard deviation	p-value (when random sampling is assumed)	Comparisons significance at the 0.05 level*	d-value 1	2	3	4
Effectiveness	1	15	2.75	0.34	0.23	none	-	0.03	0.32	0.37
	2	46	2.77	0.54			0.03	-	0.30	0.35
	3	31	2.96	0.65			0.32	0.30	-	0.06
	4	30	3.00	0.68			0.37	0.35	0.06	-

Note - Group 1: 1 – 9 employees (micro-enterprise)
    Group 2: 10 – 49 employees (small enterprise)
    Group 3: 50 – 249 employees (medium enterprise)
    Group 4: 250+ employees (large enterprise)

* Statistically significant at 0.05 level according to t-test results for independent groups
** Statistically significant at 0.01 level according to t-test results for independent groups

$^\Delta$ Medium effect in practice
$^{\Delta\Delta}$ Large effect in practice and practically significant

5.4.8 Social media platform allows for expression of sentiment and opinion

Respondents were asked to indicate their level of agreement with the following statement: This social media platform allows a customer to express sentiment and his/her opinion. The descriptive results for this statement are presented in Table 5.24. From Table 5.24 it can be seen that Facebook is the social media platform that respondents mostly agreed enabled customers/clients to express their sentiment or opinion (mean of 2.87). The second social media platform that respondents agreed enables expression of sentiment and opinion was Twitter. This information is useful to this study, because sentiment analysis and opinion mining was done on Twitter data.

Table 5.24 - Descriptive statistics regarding social media platforms allowing for expression of sentiment/opinion

Platform	N		Frequency	Percentage	Mean	Standard deviation
Facebook	119	1. Strongly disagree	17	14.29	2.87	1.03
		2. Disagree	21	17.65		
		3. Agree	42	35.29		
		4. Strongly agree	39	32.77		
Twitter	116	1. Strongly disagree	20	17.24	2.82	1.07
		2. Disagree	18	15.52		
		3. Agree	41	35.34		
		4. Strongly agree	37	31.90		
Instagram	89	1. Strongly disagree	10	11.24	2.65	0.97
		2. Disagree	32	35.96		
		3. Agree	26	29.21		
		4. Strongly agree	21	23.60		
YouTube	110	1. Strongly disagree	19	17.27	2.65	1.03
		2. Disagree	27	24.55		
		3. Agree	38	34.55		
		4. Strongly agree	26	23.64		
LinkedIn	112	1. Strongly disagree	19	16.96	2.62	1.02
		2. Disagree	30	26.79		
		3. Agree	38	33.93		
		4. Strongly agree	25	22.32		
Pinterest	104	1. Strongly disagree	21	20.19	2.54	1.08
		2. Disagree	32	30.77		
		3. Agree	25	24.04		
		4. Strongly agree	26	25.00		
Google+	107	1. Strongly disagree	21	19.63	2.43	0.98
		2. Disagree	36	33.64		
		3. Agree	33	30.84		
		4. Strongly agree	17	15.89		
Tumblr	84	1. Strongly disagree	20	23.81	2.25	0.92
		2. Disagree	30	35.71		
		3. Agree	27	32.14		
		4. Strongly agree	7	8.33		

Next, the results regarding the promotion of posts on social media platforms are presented.

5.4.9 Promotion of posts on social media platforms

Respondents were asked to indicate whether the company at which they are employed, pays to promote posts on social media platforms. Table 5.25 indicates the descriptive statistics to this question. Only 35.35% of the respondents (43 respondents) indicated that the company is paying to promote posts on social media platforms. 64.75% of the respondents (79 respondents) indicated that the company is not paying to promote posts on social media platforms.

Table 5.25 - Descriptive statistics regarding the promotion of posts

Promotion of posts on social media platforms	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Yes	43	43	35.25	35.25
2. No	79	122	64.75	100.00

The audience attracted to social media platforms also plays a role in how a company makes use of social media platforms. Results gathered regarding the audience of social media platforms are subsequently presented.

5.4.10 Audience of social media platforms

The following results identify whether respondents were of the opinion that a different social media platform attracts a different audience and how companies are determining audiences.

5.4.10.1 Different audiences for different social media platforms

Respondents were asked to give their opinion regarding different social media platforms attracting different audiences. The majority of respondents, 86.07% of the respondents, were of the opinion that each social media platform attracts its own unique audience. Only 13.93% (17 respondents) were of the opinion that different social media platforms do not attract different audiences. The descriptive statistics regarding the statement that different audiences are attracted to different social media platforms are given in Table 5.26.

Table 5.26 - Descriptive statistics regarding different audiences being attracted to different social media platforms.

Different audiences attracted by different social media platforms	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Yes	105	105	86.07	86.07
2. No	17	122	13.93	100.00

5.4.10.2 Determination of audience

Respondents were asked to indicate how the company determines the audience of their social media platforms. This open-ended question was answered by 41 respondents. The majority of respondents indicated that the company makes use of Google analytics to determine the age, location, gender and interests of customers and potential customers. Respondents also mentioned market research, market segmentation, face-to-face engagement with customers

and questionnaires/surveys are used to determine the audience that the company wants to attract to social media platforms.

The following is a response that carried meaning to the question:

> *We analyze the data that we gather from the social media platforms. Most of the social platforms, if not all have analysis tools that allow their clients to analyze the metrics, demographics, and other measures that allow us to make better business decisions.*

The limitations of social media platforms must also be considered. Respondents mentioned limitations of social media platforms as experienced by the company. The limitations are presented subsequently.

5.4.11 Limitations of social media platforms

Respondents were asked to indicate a limitation that they have experienced on a social media platform and how the company overcame this limitation. Only 28 respondents answered this open-ended question. The majority of respondents indicated that their customers/clients might not always have access to social media, the Internet or technology, especially customers/clients who stay in rural areas. Respondents also mentioned that unauthentic posts on social media platforms can cause harm to the company's reputation.

Another limitation that was mentioned was the fact that companies do not have face-to-face encounters with customers/clients when using social media. From all of the limitations listed, not one respondent indicated how the company has been able to overcome the mentioned limitation.

5.4.12 Management of social media platforms

The aim of the following three questions was to determine if employees were hired/appointed specifically to manage a company's social media platforms, if these employees had training or specific qualifications and if the company employed these employees in-house or made use of outsourced people.

5.4.12.1   Hiring/appointing of employees

Respondents were asked to indicate if the company has hired or appointed employees specifically to manage the company's social media platforms. The majority, 62.30% of the respondents, indicated that the company has not hired/appointed employees specifically to manage the company's social media platforms. Only 37.70% of respondents indicated that employees were specifically hired/appointed for the management of social media platforms. In

Table 5.27 the descriptive statistics regarding the hiring/appointing of employees for management of social media platforms are given.

Table 5.27 - Descriptive statistics regarding the hiring/appointing of employees

Hiring/appointing of employees for management of social media platforms	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Yes	46	46	37.70	37.70
2. No	76	122	62.30	100.00

The results regarding the training/qualifications of employees appointed/hired to manage a company's social media platforms are presented next.

5.4.12.2 Training/qualifications of employees

Respondents were asked to indicate whether employees appointed to manage social media platforms have received training or have a certain qualification. Only 27 respondents answered this open-ended question. 44.44% of the respondents (12 respondents) indicated that the employees who are managing the company's social media platforms have received training, while the other 55.56% of the respondents (15 respondents) indicated that the employees responsible had not received any training or have a related qualification. Qualifications mentioned include: public relations and communications, marketing, diploma in business administration, management of media. Respondents also mentioned that the employees who are responsible for the management of social media platforms had had training, for example, workshops on social media marketing.

5.4.12.3 In-house or outsourced

Respondents were asked to indicate whether employees who are responsible for social media platform management are employed in-house or are outsourced. Table 5.28 presents the descriptive statistics regarding the appointment of employees in-house or outsourced. The majority of respondents, 91.49% of respondents, indicated that employees who are responsible for the management of social media platforms are appointed in-house.

Table 5.28 - Descriptive statistics regarding employees appointed in-house or outsourced

Employee(s) appointed in-house or outsourced	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. In-house	43	43	91.49	91.49
2. Outsourced	4	47	8.51	100.00

In the following section, results gathered regarding a company's social media strategy are presented.

5.5   Section C – Social media strategy

The aim of this section was to determine which social media strategies companies use. Table 5.29 presents the questions in this section, the type of data that was gathered, the statistical techniques used to analyse the data, the population of data gathered from each question and in which tables or figures the results are depicted.

Table 5.29 - Section C: Statistical techniques used

Question	Type of data gathered	Statistical techniques used	N	Depicted in
1. What are the goals that the company want to achieve by using social media platforms?	Text	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	44	Discussed in §5.5.1
2. According to your opinion rate how integrated the company's social media goals are with the overall business objectives.	Ordinal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	61	Table 5.30
3. What are the steps that the company follows to encourage social media followers to become customers/clients?	Text	• Theme analysis/coding	25	Discussed in §5.5.3
4.1. Does the company make use of a framework or strategy when using social media platforms?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	122	Table 5.31
4.2.1. If you have answered Yes to question 4.1. please name the framework or strategy being used.	Text	• Theme analysis/coding	4	Discussed in §5.5.4.2
4.2.2. Rate the effectiveness of the named framework or strategy.	Ordinal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	121	Table 5.32
5. Does your company make use of a dashboard to manage the different social media platforms?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	122	Table 5.33
6. How is your company monitoring competitive companies, products or services?	Text	• Theme analysis/coding	29	Discussed in §5.5.7
7. According to your opinion is there a lack of control on the information being dispersed about the company, when using a social media platform?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	122	Table 5.34
8.1. According to your opinion should the company be using social media more often in its day-to-day operations?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	122	Table 5.35
8.2. If you have answered Yes to question 8.1. please state why you as employee feel that the company should be using social media more in its day-to-day operations.	Text	• Theme analysis/coding	20	Discussed in §5.5.8.2
8.3. Please express your level of agreement regarding the following statements on why the company does not use social media platforms more often.	Ordinal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	Population size is given for each item in Table 5.36.	Table 5.36 to Table 5.44.
8.4. Please list any other reasons, than those listed above, why the company does not make use of social media platforms more often?	Text	• Theme analysis/coding	19	Discussed in §5.5.8.4
9. How does your company handle a social media crisis? For example if one	Text	• Theme analysis/coding	14	Discussed in §5.5.9

Question	Type of data gathered	Statistical techniques used	N	Depicted in
of the social media platforms of the company is hacked.				

5.5.1 Desired goals that companies want to achieve by using social media

Respondents were asked what the goals are that the company want to achieve by using social media platforms. This open-ended question was answered by 44 respondents. Of the respondents 26 indicated that the company's main goal in using social media was to improve brand awareness. The other 18 respondents mentioned one of the following goals: to reach people in remote areas, to enlarge their customer base and to attract new and skilled staff.

Next, the integration of social media goals with business objectives is presented.

5.5.2 Integration of social media goals with business objectives

Respondents were asked to give their opinion regarding how integrated the company's social media goals are with the overall business objectives. Table 5.30 presents the descriptive statistics regarding the integration of social media goals with business objectives. 44.26% (27 respondents) indicated that the company's social media goals are integrated with business objectives. Only 16.39% (10 respondents) indicated that the company's social media goals are not integrated with business objectives. 14.75% (nine respondents) indicated that the company's social media goals are fully integrated with the business objectives.

Table 5.30 - Descriptive results regarding the integration of social media goals with business objectives

Integration of social media goals with business objectives	Frequency	Cumulative frequency	Percentage	Cumulative percentage	Mean	Standard deviation
1. Not integrated	10	10	16.39	16.39		
2. Moderately integrated	15	25	24.59	40.98	2.57	0.94
3. Integrated	27	52	44.26	85.25		
4. Fully integrated	9	61	14.75	100.00		

The steps taken to encourage social media followers to become customers/clients are presented below.

5.5.3 Steps taken to encourage social media followers to become customers/clients

Respondents were asked to indicate the steps that the company follows to encourage social media followers to become customers/clients. Only 25 respondents answered this open-ended question. Most of the respondents indicated that promotions are run offering incentives. Respondents also indicated that social media followers are invited to seminars and demonstrations. Some of the respondents mentioned that they have not been using social media long enough to know which steps should be taken.

The framework or specific social media strategy used by companies is subsequently presented.

5.5.4 Framework or strategy used by the company

The aim of the following two questions was to determine if companies are using a specific social media framework or strategy. The effectiveness of a mentioned framework/strategy was also indicated by respondents.

5.5.4.1 Usage of framework/strategy

Respondents were asked to indicate whether the company has been making use of a framework or strategy when using social media platforms. Table 5.31 presents the descriptive statistics regarding the use of a framework/strategy. Only 16.39% of the respondents (20 respondents) indicated that the company has been using a social media framework/strategy. The majority of respondents, 83.61% (102 respondents) indicated that the company was not using a social media framework/strategy.

Table 5.31 - Descriptive statistics regarding the use of a framework/strategy

Using a social media framework/strategy	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Yes	20	20	16.39	16.39
2. No	102	122	83.61	100.00

5.5.4.2 Framework/strategy being used

Respondents were asked to name the social media framework/strategy being used by the company. Only four respondents answered this open-ended question. The reason for such a low-response rate might be because only 20 respondents indicated that the company is using a framework/strategy. Respondents indicated that the company has created its own strategy or that it had adopted strategies used by industry leaders. From the number of responses to this question it can be deduced that respondents were reluctant or unable to share this information.

5.5.4.3 Effectiveness of named framework/strategy

Respondents were asked to rate the effectiveness of the named social media framework or strategy that the company has been using. Table 5.32 on page 134, presents the descriptive statistics regarding the effectiveness of the social media framework/strategy. The majority of the respondents, 52.38%, indicated that the social media framework/strategy is effective. 33.33% indicated that the current framework/strategy is not effective or moderately effective.

Table 5.32 - Effectiveness of named social media framework/strategy

Effectiveness of named framework/strategy	Frequency	Cumulative frequency	Percentage	Cumulative percentage	Mean	Standard deviation
1. Not effective	3	3	14.29	14.29		
2. Moderately effective	4	7	19.05	33.33	2.67	0.91
3. Effective	11	18	52.38	85.71		
4. Very effective	3	21	14.29	100.00		

Even though 20 respondents indicated that the company is using a framework or strategy, 21 respondents indicated the effectiveness of a framework or strategy.

5.5.5 Use of dashboard to manage social media platforms

Respondents were asked to indicate whether the company makes use of a dashboard to manage different social media platforms. 66.94% of the respondents (81 respondents) indicated that the company does not make use of a dashboard to manage their social media platforms, while 33.06% (40 respondents) indicated that the company does make use of a dashboard. In Table 5.33 the descriptive statistics regarding the use of a dashboard to manage social media platforms are presented.

Table 5.33 - Descriptive statistics regarding the use of a dashboard to manage social media platforms

Using of dashboard to manage social media platforms	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Yes	40	40	33.06	33.06
2. No	81	121	66.94	100.00

5.5.6 Monitoring competitive companies

Respondents were asked to indicate how the company is monitoring competitive companies, products or services. This open-ended question was answered by 29 respondents. The following list was compiled regarding how companies are monitoring their competitors:

- data analytics tools and software,
- feedback from customers/clients,
- direct interaction with customers/clients,
- research,
- determining market shares, and
- sentiment analysis.

5.5.7 Lack of control over information being distributed

Respondents were asked for their opinion regarding the lack of control over information being distributed about the company, when social media platforms are used. Table 5.34 presents the descriptive statistics regarding the lack of control over information being distributed on social

media platforms. The majority of respondents (78.69% - 96 respondents) are of the opinion that there is not a lack of control regarding the information being distributed on the company's social media platforms. Only 21.31% of the respondents are of the opinion that there is a lack of control over the information being distributed on the company's social media platforms.

Table 5.34 - Descriptive statistics regarding the lack of control over information being distributed

Lack of control on information being dispersed	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Yes	26	26	21.31	21.31
2. No	96	122	78.69	100.00

5.5.8   Higher frequency of social media use

The following four questions examined if companies should make use of social media platforms more often, as well as the reasons why companies are not using social media.

5.5.8.1   Day-to-day operations

Respondents were asked to provide their opinion regarding if the company should be using social media more often in its day-to-day operations. Table 5.35 presents the descriptive statistics regarding this question. 55.74% of the respondents (68 respondents) are of the opinion that the company should not be using social media in day-to-day operations more often, while 44.26% of the respondents (54 respondents) are of the opinion that the company should be using social media more often in its day-to-day operations.

Table 5.35 - Descriptive statistics regarding the use of social media in day-to-day operations

Use of social media in day-to-day operations should occur more often	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Yes	54	54	44.26	44.26
2. No	68	122	55.74	100.00

Reasons why companies should be using social media more often are subsequently discussed.

5.5.8.2   Why social media should be used more often

Respondents were asked to give their opinion on why companies should be using social media more in their day-to-day operations. Only 20 respondents answered this open-ended question. The responses mainly indicated that social media is becoming increasingly more important to disseminate information and to gauge what customers/clients think of companies. Social media is direct, quick and honest. Respondents were also of the opinion that social media could be used more to communicate with customers/clients. Respondents also indicated that social media can be used more by the company for marketing. Next, the results gathered regarding the reasons for not using social media platforms are presented.

5.5.8.3 Reasons for not using social media platforms

Respondents were asked to express their level of agreement regarding the following statements on why the company does not use social media more often. Table 5.36 presents the reason, the variable, the size of the population, frequency, percentage, mean and standard deviation.

Table 5.36 - Descriptive statistics regarding the reasons for not using social media platforms

Use	Variable	N		Frequency	Percentage	Mean	Standard deviation
Not enough time.	C_8_3_1		1. Strongly disagree	22	20.00	2.65	1.04
			2. Disagree	24	21.82		
			3. Agree	37	33.64		
			4. Strongly agree	27	24.55		
Not useful for the company/organization.	C_8_3_3		1. Strongly disagree	19	18.10	2.53	1.03
			2. Disagree	30	28.75		
			3. Agree	35	33.33		
			4. Strongly agree	21	20.00		
Not enough resources.	C_8_3_2	80	1. Strongly disagree	21	20.59	2.48	1.07
			2. Disagree	36	35.29		
			3. Agree	20	19.61		
			4. Strongly agree	25	24.51		
Insufficient knowledge.	C_8_3_4		1. Strongly disagree	33	30.84	2.40	1.13
			2. Disagree	24	22.43		
			3. Agree	24	22.43		
			4. Strongly agree	26	24.30		

Table 5.37 - Eigenvalues for variables C_8_3_1 to C_8_3_4

	Variable	Eigenvalue	Proportion
Not_enough_time	C_8_3_1	1.765	44.13%
	C_8_3_3	0.973	24.32%
	C_8_3_2	0.733	18.32%
	C_8_3_4	0.529	13.22%

In Table 5.37, the variable, eigenvalue of the variable as well as proportion are given. The 5-item scale has an eigenvalue that indicates the extent of variation in the items accounted for by each variable. For example, the first variable has an eigenvalue of 1.765 meaning it accounts for 44.13% of the variance (1.765/4 = 0.4413). A factor with an eigenvalue ≥ 1 explains more variance than a single observed variable. This indicated one factor, which was identified as effective.

The Cronbach alpha obtained indicated questionable validity and reliability of the factor, as can be viewed in Table 5.38 (Gliem & Gliem, 2003:87).

Table 5.38 - Cronbach alpha obtained for effectiveness construct

Factor	N	Mean	Standard deviation	Cronbach alpha
Not_enough_time	80	2.53	0.71	0.54

Table 5.39 indicates the MSA (Measurement of Sampling Adequacy) for each of the variables. The communality estimates for each variable are also presented, which indicate the percentage of variance explained by the identified factor. The overall MSA of 0.631 also indicates good sampling adequacy. The final communality estimate total was 1.90 with the one factor.

Table 5.39 - MSA and communality for variables C_8_3_1 to C_8_3_4

Variable	MSA	Communality
C_8_3_1	0.572	0.402
C_8_3_2	0.651	0.455
C_8_3_3	0.073	0.250
C_8_3_4	0.470	0.491

A two-sample *t*-test was done using the not_enough_time factor. Two groups were identified; Group 1 indicates the banking, finance and accounting business area, while Group 2 indicates the information technology and communications business area. These two groups were chosen, because the majority of data gathered by means of questionnaires fell within these two business areas. The *t*-Test results are presented in Table 5.40. Respondents who indicated that they are employed in the information technology and communications business area are agree that the company does not have enough time for improving the company's social media presence (mean of 2.52). Respondents who indicated that they are currently employed in the banking, finance and accounting business area disagree (mean of 2.39) and are of the opinion that this is not the only factor contributing to why the company is not using social media more often.

Table 5.40 - *t*-Test done using the influencing use factor

Factor	Group	N	Mean	Standard deviation	p-value (when random sampling is assumed)	d-value
Not_enough_time	1	19	2.39	0.68	0.48	0.18
	2	64	2.52	0.74		

Note: Group 1 – Banking, finance and accounting.
    Group 2 – Information technology and communications.
* Statistically significant at 0.05 level according to t-test results for independent groups

An ANOVA was done and the effect sizes regarding the number of employees were tested against the factor identified. The results of this ANOVA are presented in Table 5.41. Four groups were identified according to the type of enterprise and number of employees. Group 1 indicated as 1 – 9 employees (micro-enterprise), Group 2 was indicated as 10 – 49 employees (small enterprise), Group 3 indicated 50 – 249 employees (medium enterprise) while Group 4 indicated 250+ employees (large enterprise).

From Table 5.41 results indicate that there is a significant difference between micro-enterprises and medium enterprises regarding the opinion of not having enough time to make use of social media platforms. This is a large effect in practice. There is also a significant difference

between micro-enterprises and large enterprises, and results indicate a medium effect in practice. Respondents employed in a micro-enterprise agreed that the company does not have enough time to use social media platforms (mean of 2.89). Respondents employed in either medium or large enterprises disagreed that not having enough time is the main reason the company is not making use of social media platforms.

Table 5.41 - ANOVA results regarding the type of enterprise measured against the influence of not using social media

Construct	Group	N	Mean	Standard deviation:	p-value (when random sampling is assumed)	Comparisons significance at the 0.05 level*	d-value			
							1	2	3	4
Not_enough_time	1	15	2.89	0.65	0.03*	1 – 3 / 1 - 4	-	-0.38	-0.82 ᐃ ᐃ	-0.66 ᐃ
	2	45	2.65	0.66			-0.38	-	-0.46	-0.33
	3	31	2.32	0.70			-0.82 ᐃ ᐃ	-0.46	-	0.10
	4	29	2.39	0.76			-0.66 ᐃ	-0.33	0.10	-

Note - Group 1: 1 – 9 employees (micro-enterprise)
Group 2: 10 – 49 employees (small enterprise)
Group 3: 50 – 249 employees (medium enterprise)
Group 4: 250+ employees (large enterprise)

* Statistically significant at 0.05 level according to t-test results for independent groups
** Statistically significant at 0.01 level according to t-test results for independent groups

ᐃ Medium effect in practice
ᐃ ᐃ Large effect in practice and practically significant

Table 5.42 - ANOVA results regarding the opinion of employees on the usefulness of social media for the company

Construct	Group	N	Mean	Standard deviation	p-value (when random sampling is assumed)	Comparisons significance at the 0.05 level*	d-value			
							1	2	3	4
Usefulness_of social_media	1	15	2.33	0.82	0.15	1 – 2/ 1 – 3/1 - 4	-	0.67 ᐃ	0.56 ᐃ	0.61 ᐃ
	2	46	2.91	0.86			0.67 ᐃ	-	-0.08	0.02
	3	31	2.84	0.90			0.56 ᐃ	-0.08	-	0.10
	4	30	2.93	0.98			0.61 ᐃ	0.02	0.10	-

Note - Group 1: 1 – 9 employees (micro-enterprise)
Group 2: 10 – 49 employees (small enterprise)
Group 3: 50 – 249 employees (medium enterprise)
Group 4: 250+ employees (large enterprise)

* Statistically significant at 0.05 level according to t-test results for independent groups
** Statistically significant at 0.01 level according to t-test results for independent groups

ᐃ Medium effect in practice
ᐃ ᐃ Large effect in practice and practically significant

An ANOVA was done measuring the four groups against the opinion of employees on the usefulness of social media for the company. The results are presented in Table 5.42. From the results presented in Table 5.42 there are differences regarding the opinion of usefulness of social media for the company between the identified groups. A medium effect can be observed

in the opinion between micro-enterprises and small enterprises. A difference in the opinion of micro-enterprises and medium enterprises are also presented as a medium effect in practice. A medium effect can also be observed in the opinion between micro-enterprises and large enterprises. Respondents employed in a micro-enterprise is of the opinion that social media is not useful for the company. Respondents employed in either a small, medium or large enterprise is of the opinion that social media is useful for the company.

An ANOVA was done measuring the four groups against the perspective of the company regarding the awareness of customers. The results are presented in Table 5.43. A difference between small enterprises and large enterprises can be observed regarding the respondents opinion of the awareness of their customers regarding the company's social media presence. Respondents employed in a small enterprise is of the opinion that the company's customers are moderately aware of the company's social media presence.

Respondents employed in a large enterprise is of the opinion that their customers are aware of the company's social media presence. A medium effect in practice can be observed between these two enterprises. An assumption that can be made is that large companies have other strategies which are being used to determine the awareness of their customers or clients and are therefore of the opinion that their customers or clients are aware of the company's social media presence.

Table 5.43 - ANOVA results regarding the awareness of customers

Construct:	Group:	N	Mean:	Standard deviation:	p-value (when random sampling is assumed)	Comparisons significance at the 0.05 level*	d-value			
							1	2	3	4
Awareness_of_customers	1	15	2.60	0.91	0.11	2 - 4	-	-0.20	-0.13	0.33
	2	46	2.41	0.93			-0.20	-	0.08	0.52 △
	3	31	2.48	0.93			-0.13	0.08	-	0.45
	4	30	2.90	0.66			0.33	0.52 △	0.45	-

Note - Group 1: 1 – 9 employees (micro-enterprise)
    Group 2: 10 – 49 employees (small enterprise)
    Group 3: 50 – 249 employees (medium enterprise)
    Group 4: 250+ employees (large enterprise)

* Statistically significant at 0.05 level according to t-test results for independent groups
** Statistically significant at 0.01 level according to t-test results for independent groups

△ Medium effect in practice
△△ Large effect in practice and practically significant

An ANOVA was done measuring the four groups against how often content is posted on social media platforms. The results are presented in Table 5.44. From the results presented in Table 5.44 a difference can be observed between micro-enterprises and medium enterprises and the number of times content is posted on social media platforms. A medium effect in practice

between these two enterprises can be observed. From the results presented in Table 5.44, respondents who are employed at a micro-enterprise indicated that the company is posting content on social media platforms daily. Respondents who are employed at a medium enterprise indicated that content is posted weekly on social media platforms. An assumption that can be made is that micro-enterprises post content more often to try and grow their customer base and to become a small enterprise in the nearby future.

Table 5.44 - ANOVA results regarding how often content is posted on social media platforms

Construct	Group	N	Mean	Standard deviation	p-value (when random sampling is assumed)	Comparisons significance at the 0.05 level*	d-value			
							1	2	3	4
Content_posted	1	15	3.33	1.29	0.23	1 - 3	-	-0.22	-0.56 $^\Delta$	-0.39
	2	46	3.04	1.30			-0.22	-	-0.33	-0.16
	3	31	2.61	1.09			-0.56 $^\Delta$	-0.33	-	0.19
	4	30	2.83	1.15			-0.39	-0.16	-0.16	-

Note - Group 1: 1 – 9 employees (micro-enterprise)
Group 2: 10 – 49 employees (small enterprise)
Group 3: 50 – 249 employees (medium enterprise)
Group 4: 250+ employees (large enterprise)

* Statistically significant at 0.05 level according to t-test results for independent groups
** Statistically significant at 0.01 level according to t-test results for independent groups

$^\Delta$ Medium effect in practice
$^{\Delta\Delta}$ Large effect in practice and practically significant

Respondents were asked to list other reasons why the company does not make use of social media platforms. These results are presented next.

5.5.8.4   Other reasons why the company does not make use of social media platforms

Respondents were asked to list reasons (not included in the question discussed in §5.5.8.3), why the company does not make use of social media platforms. 19 respondents answered this open-ended question. The following list was compiled from gathered responses regarding why companies do not use social media platforms:

- Keeping up with social media trends can often distract a company from the core function of business,
- offering good products and services is better marketing than any marketing done on social media platforms,
- security is not efficient enough, which can lead to being hacked and reputational damage,
- still following traditional ways,
- reducing productivity, and
- return-on-investment (ROI) is too little.

The way companies manage social media crisis are presented next.

5.5.9 Social media crisis

Respondents were asked to indicate how the company handles a social media crisis. Only 14 respondents answered this open-ended question. From the 14 respondents, 11 respondents mentioned that when a social media crisis occurs a public apology/media statement will be placed and the problem will then be resolved. Another response was that any attempt to discredit the company is investigated and addressed and clients are informed on the other social media platforms that have not been hacked. Of the respondents, three respondents mentioned that the social media page will be taken off-line and fixed.

Next the results gathered regarding the social media metrics used by companies are presented.

5.6   Section D – Social media metrics

The aim of this section was to determine the techniques and metrics used by companies to track their social media efforts and how the company gathers analytics and social media data from the social media platforms used. Table 5.45, presents the questions asked in this section, the type of data gathered, the statistical techniques used and the population size.

Table 5.45 - Section D: Statistical techniques used

Question	Type of data gathered	Statistical techniques used	N	Depicted in
1.1. Has the company established metrics to track social media efforts to the company's objectives and goals?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage	122	Table 5.46
1.2. If you have answered Yes to question 1.1. what metrics are the company using to track social media efforts?	Text	• Theme analysis/coding	9	Discussed in §5.6.1.2
2.1. If the company has a web site, is the company monitoring and gathering analytics and data from the web site?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	122	Table 5.47
2.2. If you have answered Yes to question 2.1. how is the company gathering analytics and data from the web site?	Text	• Theme analysis/coding	28	Discussed in §5.6.2.2
3.1. How often is content posted on social media platforms?	Nominal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	122	Table 5.48
3.2. Please rate the effectiveness of the type of content posted on social media platforms (where applicable).	Ordinal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	Population size is given for each item in Table 5.49.	Table 5.49
4.1. According to your opinion do you agree that it is a good idea to link Twitter posts to automatically post to the company's Facebook page?	Ordinal data	• Frequency • Cumulative frequency • Percentage • Cumulative percentage • Mean • Standard deviation	122	Table 5.50
4.2. How does the company pull analytics and	Text	• Theme analysis/coding	13	Discussed in

Question	Type of data gathered	Statistical techniques used	N	Depicted in
data from Twitter?				§5.6.4.2
4.3. How does the company pull analytics and data from Facebook?	Text	• Theme analysis/coding	15	Discussed in §5.6.4.3
5. How does the company measure social return on investment (ROI)?	Text	• Theme analysis/coding	16	Discussed in §5.6.5

5.6.1 Metrics used to track social media efforts

The aim of the next two questions was to determine if the metrics used by the company tracks social media efforts according to the company's objectives and goals and what metrics the company was using to track these social media efforts.

5.6.1.1 Tracking that social media efforts meet company objectives

Respondents were asked to indicate if the company has established metrics to track social media efforts to the company's objectives and goals. Table 5.46 presents the descriptive statistics of the results gathered for this question. 96 respondents (78.69% of respondents) indicated that the company is not tracking social media efforts according to the company's objectives and goals. Only 26 respondents (21.31% of respondents) indicated that the company is measuring social media efforts according to the company's objectives and goals.

Table 5.46 - Descriptive statistics regarding the use of metrics to track social media efforts

Tracking that social media efforts meet company objectives	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Yes	26	26	21.31	21.31
2. No	96	122	78.69	100.00

The type of metrics being used, are presented next.

5.6.1.2 What type of metrics are being used

Respondents were asked to indicate what metrics the company has been using to track social media efforts. Only nine respondents answered this open-ended question. Respondents mentioned that the company checks the increase/decrease of social media followers, makes use of statistics that are provided by social media platforms, sentiment analysis, inquiries from new customers on how they heard about the company, and makes use of dashboards to determine the growth/decay of social media efforts.

5.6.2 Web site analytics

The following two questions gathered data regarding whether the company has a web site and how the company is gathering analytics and data from their web site.

5.6.2.1 Monitoring the company's web site

Respondents were asked that if the company has a web site, whether the company is gathering analytics and data from the web site. Table 5.47 presents the descriptive statistics regarding

the monitoring and gathering of analytics and data from the company's web site. The majority of respondents (57.38% - 70 respondents) indicated that the company is monitoring and gathering analytics and data from the web site. This is also an indication that 57.38% of the companies have web sites. 42.62% of the respondents (52 respondents) indicated that the company is not monitoring and gathering analytics and data from the company's web site.

Table 5.47 - Descriptive statistics regarding the monitoring and gathering of analytics and data from a company's web site

Monitoring and gathering analytics/data from web site	Frequency	Cumulative frequency	Percentage	Cumulative percentage
1. Yes	70	70	57.38	57.38
2. No	52	122	42.62	100.00

5.6.2.2   Techniques used to gather analytics and data from web sites

Respondents were asked to indicate the techniques used by the company to gather analytics and data from the company's web site. This open-ended question was answered by 28 respondents. All of the respondents indicated that the company is making use of Google analytics or the number of web site visits during a specific time frame.

5.6.3   Content and type of content posted

The following two questions gathered data regarding the type of content and the number of times content was posted on social media platforms.

5.6.3.1   How often content is posted

Respondents were asked to indicate how often the company posts content on social media platforms. Table 5.48 presents the descriptive statistics regarding the number of times content is posted by a company on social media platforms. 36.07% of the respondents (44 respondents) indicated that the company posts content on a weekly basis. 25.41% of the respondents (31 respondents) indicated that content is posted daily on the company's social media platforms.

Table 5.48 - Descriptive statistics regarding the number of times content is posted on social media platforms

Number of times content is posted	Frequency	Cumulative frequency	Percentage	Cumulative percentage	Mean	Standard deviation
1. Multiple times per day	15	15	12.30	12.30		
2. Daily	31	46	25.41	37.70		
3. Weekly	44	90	36.07	73.77	2.92	1.22
4. Every fortnight	13	103	10.66	84.43		
5. Monthly	19	122	15.57	100.00		

Next, the results regarding the effectiveness of the type of content posted is presented.

5.6.3.2   Effectiveness of type of content

Respondents were asked to rate the effectiveness of the type of content posted on social media platforms (where applicable). Table 5.49 present the descriptive statistics regarding the effectiveness of the type of content posted on social media platforms. The type of content which was most associated with was images, photo's, etc. and was found to be effective with a mean of 3.21. Another graphical form of content, video, was also indicated by respondents to be effective with a mean of 3.00. From the three types of content listed, text was least effective with (N = 105). Text was indicated as effective with a mean of 2.57.

Table 5.49 - Descriptive statistics regarding the effectiveness of the type of content posted on social media platforms

Type of content	N		Frequency	Percentage	Mean	Standard deviation
Images, photo's, etc.	114	1. Not effective	6	5.26	3.21	0.89
		2. Moderately effective	17	14.91		
		3. Effective	38	33.33		
		4. Very effective	53	46.49		
Video	111	1. Not effective	19	18.10	3.00	1.08
		2. Moderately effective	15	13.51		
		3. Effective	16	14.41		
		4. Very effective	28	25.23		
Text	105	1. Not effective	52	46.85	2.57	1.01
		2. Moderately effective	31	29.52		
		3. Effective	34	32.38		
		4. Very effective	22	20.95		

5.6.4   Gathering analytics from Facebook and Twitter

The next three questions required respondents to indicate the techniques used to pull analytics and data from Facebook and Twitter. Respondents was asked to indicate their opinion on whether it is a good idea to link Twitter posts of the company to automatically appear on the company's Facebook page.

5.6.4.1   Linking Twitter posts to Facebook

Respondents were asked for their opinion regarding the idea of linking Twitter posts to automatically post to the company's Facebook page. 13.11% of the respondents (16 respondents) strongly agreed that it is a good idea to link these two platforms while 22.95% of the respondents (28 respondents) strongly disagreed with this statement. Table 5.50 presents the descriptive statistics regarding the linking of a company's Twitter page with the company's Facebook page.

Table 5.50 - Descriptive statistics regarding the linking of a company's Twitter page with the company's Facebook page

Linking Twitter posts to Facebook	Frequency	Cumulative frequency	Percentage	Cumulative percentage	Mean	Standard deviation
1. Strongly disagree	28	28	22.95	22.958	2.33	0.97
2. Disagree	42	70	34.43	57.38		
3. Agree	36	106	29.51	86.89		
4. Strongly agree	16	122	13.11	100.00		

5.6.4.2 Analytics and data from Twitter

Respondents were asked to indicate how the company is pulling analytics and data from Twitter. Only 13 respondents answered this open-ended question. Techniques mentioned by respondents include: Twitter's API, Twitter's analytics web site, Twitter reporting tools, Twitter's dashboard and Hootsuite.

5.6.4.3 Analytics and data from Facebook

Respondents were asked to indicate how the company is pulling analytics and data from their Facebook page. Only 15 respondents answered this open-ended question. The techniques mentioned by respondents include: Hootsuite, Facebook advertisement manager, Facebook's reporting API and Facebook reports.

5.6.5 Social return on investment (ROI)

Respondents were asked to indicate how the company is measuring social return on investment (ROI). Only 16 respondents answered this open-ended question. Two of the respondents mentioned that the company is still in the process of determining social return on investment. The other 14 respondents indicated that the company measures the number of leads generated through social media which turned into sales, and that the company also measures the number of web site visits.

5.7 Conclusion

In this chapter, the results of the questionnaire survey based upon prior research and literature that attempts to investigate the use of social media in companies were provided. The results presented in this chapter consisted of general information of companies, social media platforms and the effectiveness of specified platforms, social media strategies and frameworks used by the company and the metrics used by companies to measure their social media efforts, and to pull analytics and collect social media data. The descriptive statistics techniques used for analysis included frequency, cumulative frequency, percentage, cumulative percentage, and mean and standard deviation. Other statistical techniques used included exploratory factor analysis, *t*-tests and ANOVA's. Random sampling was applied for these mentioned statistical techniques.

In Chapter 6, the software and the process used to analyse sentiment and opinions from Twitter data collected are described.

In Chapter 7, a discussion and conclusion to the results presented in this chapter, as well as a discussion on the opportunities for future research will be provided.

CHAPTER 6: RESULTS OF TWITTER DATA

6.1   Introduction

In this chapter, the way in which sentiment analysis and opinion mining can be done by companies today is explained. In section 6.2, the software platform, Rapidminer Studio, which was used to do sentiment analysis and opinion mining is briefly described. In section 6.3 the process used to import the Twitter data is discussed. The process built within Rapidminer Studio is discussed in section 6.4, and the results obtained from the sentiment analysis and opinion mining are discussed in section 6.5.

Utilizing social media to the benefit of companies

Chapter 1: Problem statement

Chapter 2 - Literature Study: Social media	Chapter 3 - Literature study: Mining social media data

Chapter 4: Research design

Chapter 5: Results of questionnaire	Chapter 6: Results of Twitter data
	6.1 Introduction
	6.2 Using Rapidminer Studio A description of the software platform used to do sentiment analysis and opinion mining is given in this section.
	6.3 Importing data into Rapidminer Studio The process used to import data into Rapidminer Studio is discussed in this section.
	6.4 Processes built in Rapidminer Studio Two processes were built in Rapidminer Studio. The first process made use of an operator to determine sentiment and opinion (§6.4.1) and the second process was used to aggregate the results of the sentiment analysis (§6.4.2).
	6.5 Results obtained from Twitter data Tweets of three different companies in three different business areas were gathered and the results of sentiment analysis and opinion mining is described in this section.
	6.6 Conclusion

Chapter 7: Discussion, interpretation and conclusion

Figure 6.1 - Overview of Chapter 6.

6.2   Using Rapidminer Studio

For this study, the researcher chose to make use of Rapidminer Studio to determine sentiment and opinions of customers who tweeted about their experience and their perception of a certain company. Rapidminer Studio is an open-source software platform that enables a user to make use of machine-learning techniques to perform data mining and text mining, as well as business analytics and predictive analytics (Rapidminer, 2014:19). The researcher chose this software platform because of its user-friendliness and the fact that the platform can handle great

volumes of data. Rapidminer Studio can be used by companies on a daily basis to determine sentiment and opinion of their clients/customers. Data can be imported into a previously built process and depending on the quantity of data results can be viewed graphically within minutes.

This software platform allows a user to drag and drop different operators, setting parameters and combining different operators (Rapidminer, 2014:19). By combining operators in Rapidminer Studio, a process is created. Each operator has a function and data flows between operators by connecting the output port of one operator to the input port of another operator. Rapidminer Studio continuously checks the process being built to ensure syntax conformity and gives solutions where possible problems can occur within the process (Rapidminer, 2014:19). Different data- and text-mining techniques were discussed in Chapter 3. It is necessary to understand the data and which data-mining technique should be applied within a process.

The download and installation process of Rapidminer Studio is simple. The software can be downloaded from the Rapidminer web site: http://www.rapidminer.com

Rapidminer's web site allows a user to choose from different installation packages, depending on the type of operating system installed on the user's computer. Rapidminer Studio supports all versions of Windows, Macintosh, Linux and Unix operating systems. For this study, Rapidminer Studio was installed on a computer running Windows 7 as operating system. After the user has downloaded and installed Rapidminer Studio, the user will be asked to verify the download and installation through e-mail. For this study, the researcher also identified a suitable extension for text mining and analysis. By installing this extension, the process being built is simplified.

There is a variety of extensions that can be downloaded and installed into Rapidminer Studio; extensions include web-mining extensions, Weka extension, text processing, anomaly detection, etc. Extensions can be downloaded from Rapidminer Marketplace. The user can choose to download and install an extension by using the Marketplace wizard of Rapidminer Studio or can download extensions from the following web site: https://www.marketplace.rapidminer.com

The AYLIEN Text Analysis extension was chosen for this study, because unstructured data can be used, the overall sentiment can be determined and this extension is ideal when working with blogs, reviews and social comments. After a user has installed the AYLIEN Text Analysis extension, the user will need an API key. The user can obtain the API key by activating and verifying his/her AYLIEN account by means of e-mail.

The basic principles of how a process works in Rapidminer Studio is depicted in Figure 6.2. In the rest of this chapter, each principle depicted in Figure 6.2 will be discussed.
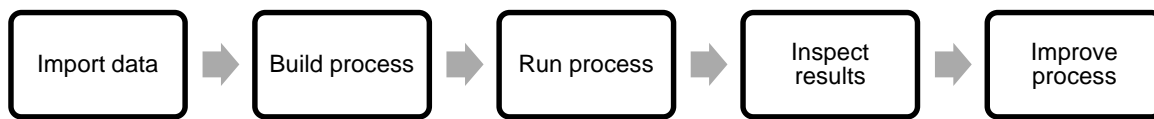
Figure 6.2 - How a process works in Rapidminer Studio

6.3   Importing data into Rapidminer Studio

Rapidminer offers an import configuration wizard, which simplifies the process of importing data. For this study Twitter data was saved in a Microsoft Excel file, which then needed to be imported. It is not necessary to make use of the import configuration wizard, but for the process used to determine sentiment and opinions, Rapidminer did not always read the data in the Microsoft Excel spreadsheet correctly. The operator used to read the Microsoft Excel file also performed better when the import configuration wizard was used. The wizard guides you through the import data process. Figure 6.3 depicts this process.
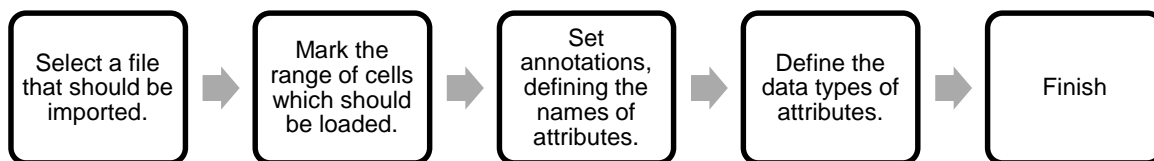
Figure 6.3 - Process for importing data into Rapidminer Studio

The process used for sentiment analysis and opinion mining are subsequently described.

6.4   Process built in Rapidminer Studio

Two processes were built to determine sentiment and opinions. The first process analysed the data and categorised the tweet according to sentiment and opinion. The second process was used to aggregate the sentiment and to test the precision of the sentiment analysis by using three classification models: a decision tree, Naïve Bayes classifier and a Support Vector Machine (SVM).

6.4.1   Process used to determine sentiment and opinions

The tweets were analyzed according to sentiment and opinion. Sentiment was classified as positive, neutral or negative. The tweet was also used to determine the opinion, which could be categorized as objective or subjective. Two operators were used to determine the sentiment and opinions. Figure 6.4 is a representation of the process discussed in this section.

A description of the operators is as follows:

1. Read Excel: This operator was used to load data from the Microsoft Excel spreadsheet. It should be specified which spreadsheet of the workbook should be used. Each row represented a tweet, and the spreadsheet contained one column, which represented the attribute, and tweet text.

2. Analyze Sentiment: This operator is only available when the AYLIEN Text Analysis extension is installed. It allows a user to extract sentiment from text, for this study a tweet, and provides the user with valuable insight into the author's, for this study customer's, emotions and perspective. The sentiment of the tweet was classified as positive, neutral or negative. Opinions were classified as subjective (a reflection of the customer's opinion) or objective (an expression of fact).

Figure 6.4 - Process used to determine sentiment and opinions

6.4.2 Process used to aggregate sentiment

After the sentiment of each tweet had been determined, the tweets, as well as the sentiment of each tweet were processed again to test the precision. The main process consisted of five operators. This process was only used on tweets classified as positive or negative during sentiment analysis.

The operators and a description of each operator are as follows:

1. Read Excel: This operator was used to load data from the Microsoft Excel spreadsheet. Each row represented a tweet, and the spreadsheet contained two columns, which represented the attributes, sentiment and tweet text.

2. Process Documents to Data: This operator was used to generate word vectors from string attributes, and consisted of a sub-process. The sub-process is depicted in Figure 6.5 and consisted of the following operators:

   2.1 Transform Cases: This operator transforms all characters in a document to either lowercase or uppercase.

   2.2 Tokenize: The function of this operator is to split the text of a document into a sequence of tokens and the user can specify what the splitting points must be. The best option is to create tokens that consist of a single word.

   2.3 Filter Stopwords: This operator has a built-in stopword list and filters English stopwords from a document by removing every token.

2.4 Stem (Porter): This operator has a built-in stemming algorithm that applies an iterative, rule-based replacement of word suffixes, reducing the length of words.
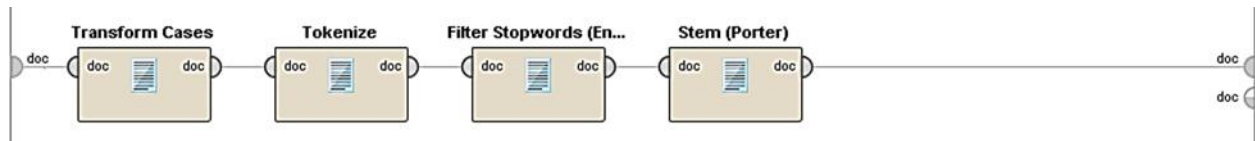
Figure 6.5 - Process Documents to Data sub-process

3. Set Role: This operator is usually used to change the role of one or more attributes of the input. For this study, this operator was used to set the right roles for attributes before applying the desired operator.

4. Select Attributes: This operator was used to select which attributes of the input should be kept and which attributes should be removed. This operator is especially useful when the user has not pre-processed the data on a Microsoft Excel spreadsheet. This operator contributes to the selection of required attributes.

5. Validation: The Split Validation operator is a nested operator, consisting of two sub-processes: a training sub-process and a testing sub-process. The training sub-process was used for learning or building the model. The trained model was then applied in the testing sub-process. The performance of the model was also measured during the testing phase. The training and testing sub-processes are depicted in Figure 6.6 and consisted of the following operators:

5.1 Decision Tree/Naïve Bayes/Support Vector Machine (SVM): These classification models were used to test the precision of the sentiment analysis and as prediction guideline.

5.2 Apply Model: This operator applies an already learnt or trained model on the given input.

5.3 Performance: This operator is used as a statistical performance evaluation of the classification models used.
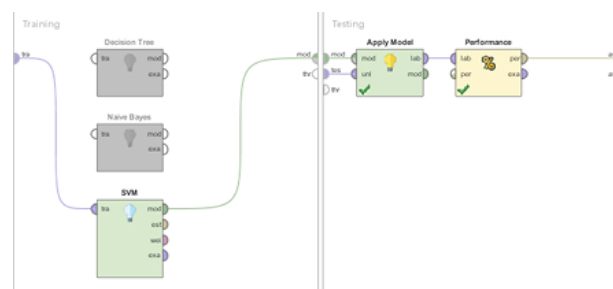
Figure 6.6 - Validation sub-processes

Figure 6.7 is a representation of the process discussed in this section.
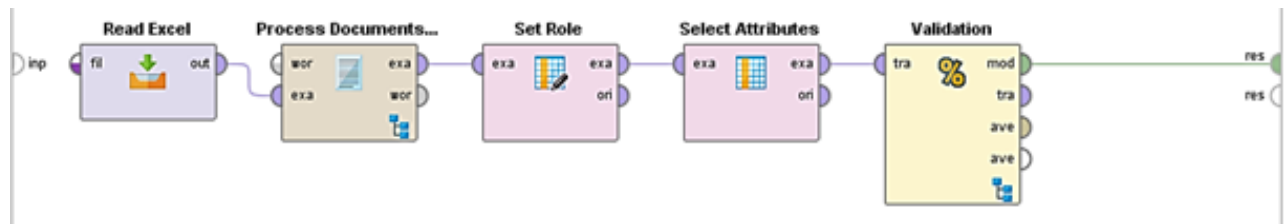
Figure 6.7 - Process used to aggregate sentiment

6.5 Results obtained from Twitter data

Before data collection commenced, three companies in different business areas were chosen, namely retail, banking and insurance. In the past, data mining has been applied successfully in these business areas. The choice of these business areas was also influenced by the literature as discussed in Chapter 3, section 3.2.1. To recapitulate a figure discussed in Chapter 3 (Figure 3.2), these business areas can also be divided into different segments. For this study, the focus was on customer retention. Companies today want to understand their clients'/customers' needs and this can be achieved through sentiment analysis and opinion mining. The chosen companies' names will be kept anonymous during the discussion of results and will be referred to as Company X, Company Y and Company Z.

6.5.1 Company X

Company X is a well-known international retail company. This company is a large enterprise with approximately 9 000 employees in South Africa alone. This company's mission is to provide happiness, inspire moments and make a difference in communities. Company X has a rich history, which started in 1886, in Atlanta, Georgia and was first shipped to, and distributed and sold in South Africa in the 1930's. The company has 25 brands that are currently being sold across South Africa.

5 994 tweets concerning Company X were gathered and analyzed. All tweets were gathered within a timeframe of a month. All duplicate tweets were removed. The AYLIEN Text Analysis extension returns five attributes of each tweet as results: the polarity, the polarity confidence, the subjectivity, the subjectivity confidence and the tweet text. The polarity indicates the sentiment, while the subjectivity indicates the opinion. The polarity confidence and subjectivity confidence range from 0 – 1, and a value close to 1 indicates a higher confidence level. Table 6.1 presents the sentiment and the average confidence levels of polarity for Company X. Only 17.60% of the tweets concerning Company X were positive, while 75.48% were neutral and 6.92% were negative.

Table 6.1 - Average confidence levels of polarity for Company X

Sentiment	Number of tweets classified	Percentage of overall tweets	Average polarity confidence	Overall average of polarity
Positive	1055	17.60	0.72	
Neutral	4524	75.48	0.80	0.77
Negative	415	6.92	0.67	

42.53% of the tweets were classified as subjective meaning that an opinion was tweeted while 57.47% were objective meaning that a fact was tweeted. The results regarding the average confidence levels of subjectivity levels are presented in Table 6.2.

Table 6.2 - Average confidence levels of subjectivity for Company X

Opinion	Number of tweets classified	Percentage of overall tweets	Average subjectivity confidence	Overall average of subjectivity
Subjective	2549	42.53	0.97	0.97
Objective	3445	57.47	0.97	

Only the positive and negative tweets were used when aggregating the sentiment. Only 1 470 tweets concerning Company X were classified as either positive or negative. 441 tweets were used as a sample of the process described in section 6.4.2. Rapidminer Studio and the process built to aggregate sentiment determines the sample size based on the number of tweets gathered, as well as the sample size that will be used for validation and testing. A Support Vector Machine (SVM) was applied within the process. The accuracy of the model was 79.14%, indicating 20.86% error in the model, which can be reduced. 45 tweets that were actually positive were predicted as positive, and six tweets that were actually negative were predicted as positive. 86 tweets that were actually positive were predicted as negative and 304 tweets that were actually negative were predicted to be negative. The results regarding the precision of the sentiment determined for Company X are presented in Table 6.3.

Table 6.3 - Precision of sentiment determined for Company X

	True positive	True negative	Class precision
Pred. positive:	45	6	88.24%
Pred. negative:	86	304	77.95%
Class recall:	34.35%	98.06%	

The sentiment and opinions analyzed regarding Company Y are discussed next.

6.5.2 Company Y

Company Y is a South-African company specializing in banking and providing financial services to customers/clients. The history of this company dates back to 1838. Company Y forms part of the big four commercial banks in South Africa. This company does not only serve South Africa, but is currently expanding to neighbouring countries, such as Botswana.

620 tweets concerning Company Y were gathered within a timeframe of a month. All duplicate tweets were removed. Only 27.42% of the tweets were classified as positive, while 41.77% were classified as neutral and 30.81% were classified as negative. The average confidence levels of polarity for Company Y are presented in Table 6.4.

Table 6.4 - Average confidence levels of polarity for Company Y

Sentiment	Number of tweets classified	Percentage of overall tweets	Average polarity confidence	Overall average of polarity
Positive	170	27.42	0.82	
Neutral	259	41.77	0.71	0.76
Negative	191	30.81	0.76	

53.39% of Company Y's tweets were subjective (an opinion regarding the company), while 46.61% were classified as objective (a fact stated regarding the company). The average confidence levels of subjectivity for Company Y are presented in Table 6.5.

Table 6.5 - Average confidence levels of subjectivity for Company Y

Opinion	Number of tweets classified	Percentage of overall tweets	Average subjectivity confidence	Overall average of subjectivity
Subjective	331	53.39	0.99	
Objective	289	46.61	0.98	0.98

To test the precision of the sentiment analysis, a decision tree was applied within the process described in §6.4.2 and an example of the results for Company Y's tweets is given below. Different words are identified in the tweet. From the results below an interpretation of the word 'great' can be made according to the TF-IDF score. This score is used to rank the importance of words, as well as the frequency of a word. If the TF-IDF score for the word 'great' is greater than 0.081 then a tweet with the word 'great' in was classified as positive and if the TF-IDF score is smaller than 0.081 the tweet was classified as negative. The classification was done similarly for the remaining words listed below.

```
|   great > 0.081: positive {positive=10, negative=0}
|   great ≤ 0.081
|   |   breez > 0.127: positive {positive=6, negative=0}
|   |   breez ≤ 0.127
|   |   |   win > 0.117: positive {positive=6, negative=0}
|   |   |   win ≤ 0.117
|   |   |   |   honour > 0.126: positive {positive=5, negative=0}
|   |   |   |   honour ≤ 0.126
|   |   |   |   |   welcom > 0.135: positive {positive=4, negative=0}
|   |   |   |   |   welcom ≤ 0.135
|   |   |   |   |   |   date > 0.326: positive {positive=3, negative=0}
|   |   |   |   |   |   date ≤ 0.326
|   |   |   |   |   |   |   fun > 0.139: positive {positive=3, negative=0}
|   |   |   |   |   |   |   fun ≤ 0.139
|   |   |   |   |   |   |   |   fyi > 0.144: positive {positive=3, negative=0}
|   |   |   |   |   |   |   |   fyi ≤ 0.144
|   |   |   |   |   |   |   |   |   start > 0.127: positive {positive=3, negative=0}
|   |   |   |   |   |   |   |   |   start ≤ 0.127
|   |   |   |   |   |   |   |   |   |   absolut > 0.160: positive {positive=2, negative=0}
|   |   |   |   |   |   |   |   |   |   absolut ≤ 0.160
|   |   |   |   |   |   |   |   |   |   |   afta > 0.152: positive {positive=2, negative=0}
|   |   |   |   |   |   |   |   |   |   |   afta ≤ 0.152
|   |   |   |   |   |   |   |   |   |   |   |   amaz > 0.223: positive {positive=2, negative=0}
|   |   |   |   |   |   |   |   |   |   |   |   amaz ≤ 0.223
|   |   |   |   |   |   |   |   |   |   |   |   |   awesom > 0.120: positive {positive=2,
negative=0}
|   |   |   |   |   |   |   |   |   |   |   |   |   awesom ≤ 0.120
|   |   |   |   |   |   |   |   |   |   |   |   |   |   bainbridg > 0.114: positive {positive=2,
negative=0}
```

153

```
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   bainbridg ≤ 0.114
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |     celebr > 0.117: positive {positive=2,
negative=0}
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |     celebr ≤ 0.117
```

The accuracy of the model was 85.82%, indicating 14.18% error in the model, which can be reduced. The Validation operator makes use of random sampling when choosing tweets to train and test through the model. From the 361 tweets classified either as positive or negative, a sample size of 261 was used. Rapidminer Studio and the process built to aggregate the sentiment determines the sample size of the data sets for validation and testing. 92 tweets that were actually positive were predicted as positive. Six tweets that were actually negative were predicted as positive. 31 tweets that were actually positive were predicted as negative and 132 tweets that were actually negative were predicted to be negative. The precision of sentiment determined for Company Y is presented in Table 6.6.

Table 6.6 - Precision of sentiment determined for Company Y

	True positive	*True negative*	*Class precision*
Pred. positive	92	6	93.88%
Pred. negative	31	132	80.98%
Class recall	74.80%	95.65%	

The results of sentiment and opinion regarding Company Z will be discussed next.

6.5.3   Company Z

Company Z specializes in short- and long-term insurance. This company was founded in 1998 and has expanded to neighbouring countries, such as Namibia. This company also has subsidiaries in Australia and New Zealand. Only 11 tweets concerning Company Z were gathered within a timeframe of a month. This small number of tweets gathered can also indicate the company's involvement and use of social media, as well as whether the company responds to social media posts and tweets. There were no duplicate tweets. 81.82% of the tweets were classified as positive, while 18.18% were classified as negative. The average confidence levels of polarity for Company Z are presented in Table 6.7.

Table 6.7 - Average confidence levels of polarity for Company Z

Sentiment	*Number of tweets classified*	*Percentage of overall tweets*	*Average polarity confidence*	*Overall average of polarity*
Positive	9	81.82	0.91	0.89
Negative	2	18.18	0.80	

Table 6.8 presents the average confidence levels of subjectivity for Company Z. 45.45% of the tweets were classified as subjective (stating an opinion regarding the company) while 54.55% were classified as objective (stating a fact regarding the company).

Table 6.8 - Average confidence levels of subjectivity for Company Z

Opinion	Number of tweets classified	Percentage of overall tweets	Average subjectivity confidence	Overall average of subjectivity
Subjective	5	45.45	0.99	0.99
Objective	6	54.55	0.98	

The precision of the sentiment analysis was tested by applying a decision tree, as well as a Support Vector Machine (SVM) within the process described in §6.4.2. Both of these classification models delivered the same results as presented in Table 6.9. The accuracy of the model was 75.00%, indicating 25.00% error in the model, which can be reduced. From the 11 tweets classified either as positive or negative, a sample size of four was used. Three tweets that were actually positive were predicted as positive, and one tweet that was actually negative was predicted as positive.

Table 6.9 - Precision of sentiment determined for Company Z

	True positive	True negative	Class precision
Pred. positive	3	1	75.00%
Pred. negative	0	0	0.00%
Class recall	100.00%	0.00%	

6.6   Conclusion

In this chapter, the results regarding Twitter data for three companies, each in a different business area were provided. The process followed to do sentiment analysis and opinion mining was elaborated on. By using a software platform, such as Rapidminer Studio, a company can easily determine the sentiment and opinions of customers/clients. The processes built for this study can be improved according to the needs of the company regarding gathered social media data. Sentiment was classified as positive, neutral or negative and opinions were classified as either subjective or objective.

Sentiment was indicated as polarity while an opinion was indicated as subjectivity. The level of confidence for both polarity and subjectivity was also determined to ensure that, for example a negative tweet, was truly classified as negative. The precision of the determined sentiment was also tested by using decision trees, a Naïve Bayes classifier or a Support Vector Machine (SVM). The Naïve Bayes classifier was used for each precision test done, but the accuracy of the model was often above 75.00%. The decision tree, as well as the Support Vector Machine (SVM) models that were applied had an accuracy level of more than 75.00% for each precision test done. The results of the model with the highest accuracy were used.

Rapidminer Studio offers great opportunities for analyzing sentiment and opinions. A discussion and conclusion regarding the results of Chapter 5 and Chapter 6 follow.

CHAPTER 7: DISCUSSION AND CONCLUSION

7.1 Introduction

The aim of this chapter is to provide a discussion and conclusion regarding the survey results presented in Chapter 5, as well as the results of the analysis of the Twitter data in Chapter 6. In section 7.2, a discussion regarding the analysed results of the survey is given. Section 7.2 is divided into the four sections of the questionnaire: general information, social media platforms, social media strategy and social media metrics. In section 7.3, a discussion regarding the results obtained from analysed Twitter data is given. The contributions of this study are discussed in section 7.4. The answer to the research question of this study and an explanation of how the aims and objectives defined for this study were met is given in section 7.5. In section 7.6, the limitations of this study are pointed out, followed by a discussion of future work in section 7.7. An overview of this chapter is given in Figure 7.1.

Utilizing social media to the benefit of companies

Chapter 1: Problem statement

Chapter 2 - Literature Study: Social media	Chapter 3 - Literature study: Mining social media data

Chapter 4: Research design

Chapter 5: Results of questionnaire	Chapter 6: Results of Twitter data

Chapter 7: Discussion, interpretation and conclusion

7.1 Introduction

7.2 Discussion regarding survey results
A discussion regarding the survey results are given. Results for each section of the questionnaire are discussed.

7.3 Discussion regarding results obtained from Twitter
A discussion regarding the software platform used to analyze tweets and the results obtained regarding this process are discussed.

7.4 Contribution of this study
The contributions made by this study are discussed as well as the basic framework created.

7.5 Answering the research question and achieving secondary objectives
The research question is answered in this section. Nine different secondary objectives were set at the beginning of this study, in this section the achievement of these secondary objectives are discussed.

7.6 Limitations of this study

7.7 Future work and research
Suggestions for future work and research are discussed in this section.

7.8 Conclusion

Figure 7.1 - Overview of Chapter 7

7.2 Discussion regarding survey results

In section 7.2.1, general information of the companies who participated in this study is discussed. Results gathered regarding social media platforms are discussed in section 7.2.2. The types of social media strategy followed by companies today are discussed in section 7.2.3. In section 7.2.4, different techniques used by companies to measure their social media performance are discussed.

7.2.1 Discussion of Section A – General information

The data analyses indicated that the majority of respondents were mainly employed in information technology and communications (53.28%), followed by banking, finance and accounting (16.39%), as well as retail, sales and marketing (14.75%). From the literature investigated, data mining has been applied successfully in the retail, banking and insurance sectors. This was of importance for this study as text- and data-mining techniques were applied on social media data gathered for each of these business areas. Only 4.10% of the respondents indicated that they are employed in the legal business area.

Most companies have not identified techniques and tools that could be used to improve their social media presence and to improve customer relationship management. The fact that most of the companies that participated in this study are in the information technology and communications industry was surprising as no new technology was being used to measure the company's social media return.

The majority of respondents (37.70%) indicated that the company at which they are employed can be categorized as a small enterprise consisting of 10 – 49 employees. 25.41% of the respondents were employed in a medium-sized enterprise consisting of 50 – 249 employees, followed by 24.59% of the respondents being employed in large enterprises comprising 250+ employees. Only 12.30% of the respondents indicated that they were employed in micro-enterprises consisting of 1 – 9 employees. The size of the company can also play a role when using social media as larger companies can assign more employees to manage social media platforms and to investigate techniques that can increase their number of followers on social media networks.

Of the respondents 45.08% were of the opinion that social media is useful for the company. The term usefulness of social media can be defined differently by each respondent, for example the usefulness of social media can be that social media increase sales, builds the company's reputation, improves marketing of certain products or services, etc. Only 9.84% of the respondents were of the opinion that social media is not useful to the company.

Respondents with the opinion that social media is not useful for the company could have developed this opinion because, for instance the company had experienced a social media crisis. Respondents with the opinion that social media is not useful for the company could also have developed this opinion if the company did not have the correct strategy in place when using social media and that the company is currently not gaining from the use of social media platforms. A discussion regarding the survey results gathered for social media platforms follows.

7.2.2 Discussion of Section B – Social media platforms

Companies should tread lightly when allowing their employees to visit social media platforms at work, as this can have an influence on an employee's work performance. Of the respondents 81.15% indicated that they are allowed to visit social media platforms during working hours. An employee can cause a company great harm when posting harmful content that can be connected with the company, which in turn can have an influence on the company's reputation, as well as on sales of products or services.

The use of social media in companies is still in its infant years as most of the respondents (29.51%) indicated that the company had only been using social media for a period of one to two years. This was followed by 27.87% of the respondents that indicated that company has only been using social media for a period of 9 to 12 months. Only 17.21% of the respondents indicated that the company had been using social media for five+ years. This also indicates that the use of social media in companies is still a new topic of research.

From eight different social media platforms (Facebook, Twitter, LinkedIn, Google+, YouTube, Pinterest, Instagram and Tumblr), Facebook and Twitter proved to be the most effective when used for marketing. Facebook and Twitter were also rated as the most effective when used for branding and customer relationship management. From literature that is available, these two social media platforms are mostly used by companies. Pinterest, Instagram and Tumblr proved to be less effective when used for marketing, branding and customer relationship management. Companies also make use of bulk SMS campaigns, Yammer and WhatsApp messenger to engage with customers and do marketing of products or services.

The respondents also indicated that the company regularly uses social media platforms to increase brand awareness, to advertise products or services and for business to business purposes. Social media is used moderately for gaining feedback from customers, engaging with customers, analyzing the competition in a similar industry and offering promotional offers or items. Other uses of social media included community engagement, talent recruitment and

distributing information regarding the company. A social media platform that is often found very effectively for talent recruitment and advertising job opportunities is LinkedIn. When a company decides to hire or appoint a new employee, it can also visit the future employee's social media platforms to identify other aspects of the future employee that could not have been captured during an interview.

Respondents indicated that social media platforms are effective for advertising a product or service. Social media platforms are viewed as a marketing platform and companies are only starting to realize the potential that social media platforms can have for engaging with customers or clients. Companies can perform sentiment analyses and opinion mining on social media data by using text-, data- and web-mining techniques. Respondents indicated that they agree that social media platforms, such as Facebook, Twitter, Instagram, YouTube, LinkedIn and Pinterest allow customers or clients to express their sentiment and opinion regarding a specific product, service or brand.

Very few respondents (35.25%) indicated that the company was paying to promote posts on social media platforms. This indicates that a social media post can be powerful. A post that contains specific elements, such as graphical content and a simple yet powerful message can be shared by customers or clients on different social media platforms. Customers or clients can also give their opinion regarding a post made by a company, which can lead to a chain reaction between social media users using the same product, service or brand. Companies should make use of techniques and tools to track the participation of their customers or clients. Technology today allows the tracking of customer's activities between different web sites and social media platforms.

Social media platforms also have barriers and limitations that companies do not always know how to overcome. Some of the greatest limitations listed by respondents included that the company's customers or clients do not always have access to Internet or technology. If a company's customers or clients are mainly situated in rural areas where there is no Internet or technology, a company should create opportunities in this area for their customers or clients to come into contact with social media platforms. A company can, for example place a touch screen computer in a shop which is located in a rural area where customers or clients can then experience social media and give feedback regarding products or services.

A company's perspective of the awareness of their customers or clients regarding the company's social media presence also influences the social media strategy, which a company employs. The majority of respondents (86.07%) were of the opinion that each social media

platform attracts its own unique audience. The awareness of customers or clients should be tested and evaluated continuously. Techniques currently being used by companies include Google analytics, Facebook reports and Hootsuite. Most of the respondents were of the opinion that the company's customers or clients are aware of the social media platforms being used by the company. If a company can identify its target audience and the current awareness of that audience, content being posted on social media platforms can be developed according to this identified audience.

Only 37.70% of respondents indicated that the company at which they are employed is appointing people to manage social media platforms. Employees who are being hired to manage social media platforms have qualifications either in public relations and communications, marketing, business administration or management and media. Companies are also sending employees responsible for the management of social media platforms to workshops regarding social media. The employees responsible for managing social media platforms are employed in-house and are not outsourced. The management of a company might feel that it has more control over employees responsible for the management of social media platforms if these employees are employed in-house. Next, results gathered regarding social media strategies used by companies are discussed.

7.2.3 Discussion of Section C – Social media strategy

The majority of respondents indicated that the company's business objectives and social media goals are integrated. The main goal of social media use is to increase brand awareness. Other uses listed include reaching customers or clients in remote areas, enlarging the customer base and attracting new and skilled staff. Only 16.39% of the respondents indicated that the company has a specified social media strategy or is using a social media framework.

Respondents were either not eager or not able to share the type of strategy or t framework being used by the company. This can possibly be to not allow competitive companies to get hold of a strategy and framework. Companies can monitor competitive companies by gaining feedback from customers or clients either by means of social media platforms or direct interaction, determining market shares and using data analytics tools and software to perform sentiment analysis and opinion mining.

Respondents were also either not eager or not able to share the steps that the company is taking to encourage social media followers to become customers or clients. Respondents indicated that incentives are often used to increase customer numbers. From the results gathered in this section, it became clear that companies are uncertain about the steps that

need to be taken to ensure that social media followers become customers or clients. The type of company and the type of customer should be considered when developing a strategy or deciding on a framework, as each company is unique. Of the respondents, 33.33% indicated that the strategy or framework currently being used by the company is either not effective or moderately effective.

Only 33.06% of the respondents indicated that the company is making use of a dashboard to manage its social media platforms. Most social media platforms offer built-in dashboards, which can be used to measure the followers' (customers') activities on a company's social media page. Of the respondents, 44.26% indicated that the company should be using social media more in its day-to-day operations while 55.74% of the respondents are of the opinion that the company should not be using social media more in its day-to-day operations. This can be because of the fact that companies experience a lack of control on the information being distributed, but the majority of the respondents (78.69%) are of the opinion that this does not have an effect on why the company is not using social media more often.

Four reasons were listed why companies do not use social media more often including not having enough time, that social media is not useful for the company, that the company does not have enough resources and that there is a lack of knowledge regarding social media. From these four reasons, the reason mentioned most was that the company does not have enough time to manage and improve the company's social media presence. The size of the company can also have an influence on why the company is not using social media more in its day-to-day operations.

There is a significant difference between micro-enterprises (1 – 9 employees) and medium enterprises (50 – 249 employees), as well as micro-enterprises and large enterprises (250+ employees) when it comes to the time spent on managing social media platforms. An assumption that can be made is that medium- and large enterprises can spend more time on improving their social media presence because they have more employees who can be assigned to this task. Another assumption that can be made is that micro-enterprises do not have as many resources as medium or large enterprises and do not necessarily have as much time to improve their social media presence because the employees are more focused on core business functions.

Small medium and large enterprises are of the opinion that social media is useful to the company. Micro-enterprises disagree and this might be because these enterprises do not spend as much time and effort on improving their social media presence and gaining more

customers. An assumption that can be made is that micro-enterprises choose not to use social media platforms as often as the other three enterprises because they feel there is more control in the company if the company and customer base remains a certain size.

Small enterprises are of the opinion that their customers are not as aware of their social media presence. Large enterprises are of the opinion that their customers or clients are well aware of their social media presence. This can be because large enterprises have more resources and time to manage their social media platforms thereby increasing the number of times that content can be posted on social media platforms and the time that can be spent engaging with customers or clients. Medium enterprises tend to post content on social media platforms more often than micro-enterprises. Medium companies want to increase their number of employees and their customer base to become a large enterprise.

Respondents are also of the opinion that keeping up with social media trends can often distract a company from its core business functions. If a product or service is good, it is not essential to make use of social media to market the presented product or service, the customer base will automatically grow. From results obtained in the survey, an assumption can be made that some companies today are still following traditional ways and are not interested in social media because social media may, in their opinion, reduce productivity and can lead to reputational damage. Companies can place an apology/media statement if one of the company's social media platforms is hacked. The metrics used by companies to track their social media efforts will be discussed next.

7.2.4 Discussion of Section D – Social media metrics

The majority of respondents (78.69%) indicated that the company is not using any metrics to track social media efforts. Companies that track their social media efforts use techniques, such as checking the increase or decrease of social media followers, making use of social media platform dashboards, sentiment analysis and inquiring from new customers how they heard about the company.

Companies that have a web site should make use of web mining to track and measure their customers' activity on the web site. 57.38% of the respondents indicated that the company is gathering analytics and data from its web site. Facebook and Twitter offer techniques such as Twitter's API (Application Programming Interface), Facebook's reporting API, Twitter's analytics web site, Twitter's reporting tools, dashboards, Facebook advertisement manager and Hootsuite, which can be used to gather analytics and data from these two social media platforms. The majority of respondents are of the opinion that it is not a good idea to link a

162

company's Twitter posts to the company's Facebook page. If social media platforms are not linked, a social media crisis can be better managed.

The results obtained from the survey indicate that the size of a company can also have an influence on the number of times that content is posted on social media platforms. The majority of the respondents indicated that content is posted on a weekly basis on the company's social media platforms. The type of content that is most effective and draws followers' attention rapidly is images and photo's followed by video. Customers or clients are often asked to post a photo when using a product or service on the specific company's social media page. This is used as an advertising and marketing strategy. The majority of companies have not yet identified a process to determine social return on investment. Companies are measuring the number of leads generated through social media, which turned into sales and the number of web site and social media platform visits during a specific time frame.

Next, the results obtained from analyzing tweets will be discussed.

7.3   Discussion regarding results obtained from analysis of Twitter data

For this study, the software platform Rapidminer Studio was used to analyse tweets gathered according to sentiment and opinion. Rapidminer Studio was used because building a process using this platform is easy and fast and this software platform can handle large volumes of data. When a company chooses to make use of social media data, it is recommended that a large data set should be used for processing. Different classification models can then be applied to the larger data set to ensure that sentiment and opinion are extracted more precisely. When working with a larger data set, the sentiment and opinions of customers or clients can be determined in more detail.

With a larger data set, a decision tree can be used and a word cloud can then be built from the results of the decision tree. This can be presented to management as a summary of the sentiment and opinions of customers or clients. By using a software platform, such as Rapidminer Studio, a company that has little time and a small number of resources can determine the overall sentiment and opinions from customers or clients. The process built in such a software platform will depend on the information that the company wants extracted from social media data. A company should ensure that its social media presence and engagement with customers or clients is good and that the quantity of user-generated content created is encouraged by the company. This will then ensure that a large data set can be used for analysis.

For this study, social media data of three different South African companies was gathered. The companies' names were kept anonymous and were referred to as Company X, Company Y and Company Z.

From the 5 994 tweets that were gathered and analysed according to sentiment for Company X, the majority of tweets (75.48%) were categorized as neutral. When a tweet was categorized as neutral, this meant that the customer did not express any negative or positive feelings regarding the company. The tweets that were categorised as positive (17.60%), was an indication that the customers expressed and have a positive perception regarding the products delivered by Company X. Only 6.92% of the tweets were categorised as negative for Company X. After a sample of the positive and negative tweets was placed through another process and a Support Vector Machine (SVM) was applied in the process, 86 of the tweets that were first categorized as negative were actually positive. This indicated that the perception of customers regarding Company X and the products delivered by this company was positive.

Opinions were classified as subjective (a reflection of the customer's opinion) or objective (an expression of fact). From the 5 994 tweets gathered of Company X, the majority of tweets (57.47%) were categorized as objective, while the other 42.53% were categorized as subjective reflecting the customer's opinion.

For Company Y, a total of 620 tweets were gathered and analysed. Only 41.77% of the tweets were categorized as neutral. From the remaining tweets, most of the tweets were categorized as negative (30.81%), while only 27.42% of the tweets were categorized as positive. A decision tree was applied within the second process and the tweets were already categorised quite precisely. Only 31 tweets that were actually positive were predicted as negative. From this sentiment analysis an assumption can be made that the customers of Company Y had mixed feelings, because the percentage of tweets classified as positive was almost similar to the percentage of tweets classified as negative. The majority of customers (53.39%) expressed an opinion regarding the company's products and services, while 46.61% of the customers who tweeted expressed a fact regarding the company.

Only 11 tweets concerning Company Z were gathered. Because the data set was so small, the company cannot get a general perception of how customers are reacting regarding its products or services. None of the tweets were categorized as neutral when determining the sentiment. The majority of tweets (81.82%) were categorized as positive. Only 18.18% of the tweets were categorized as negative. The majority of customers (54.55%) tweeted a fact concerning

Company Z and the company's products or services. Only 45.45% of the customers tweeted their own opinion regarding Company Z.

These three different companies and business areas were chosen because from available literature, data-mining techniques have already been successfully applied in these three different business areas. Companies can use the results obtained from sentiment analysis and opinion mining to identify negative feelings and perceptions regarding the company, products or services of the company. Negative tweets can be investigated and the negativity regarding a product or service can be addressed.

Next, the contributions made by this study are discussed.

7.4   Contribution of this study

This study and research contributed to the following areas:

- Improve knowledge regarding the value of social media if managed correctly.

The literature studied, as well as the results obtained from the survey and the analysis of Twitter data can improve the knowledge not only of companies and business owners, but social media users and social media platform developers.

- Provide companies with a basic framework and available techniques that can be used to analyse customers'/clients' feelings and paradigms about the company, which can then be used to improve customer relationship management.

A basic framework was created in which possible techniques are mentioned, which can be used by companies to determine sentiment and opinions from social media data. This framework also included techniques, which can be used to manage a company's social media platforms. The framework can serve as a guideline or tool for companies that are struggling with the effective use of social media platforms and are not finding social media platforms to be useful for the company. The framework consists of 11 guidelines that can be followed. The frameworks discussed in Chapter 2, section 2.5 were used as basis during the development of this framework and were altered according to available literature, results obtained from the survey and knowledge gained from this study. Table 7.1 presents each guideline of the developed framework and from which source the guideline was developed. The framework is depicted in Figure 7.2 on page 166.

Table 7.1 - Development of a basic framework

Guidelines	Developed according to
1. Align the company's social media goals with the company's business objectives.	Survey results
2. Prioritize social media goals.	Framework created by Jeffrey (2013) and Paine (2015)
3. Identify a timeframe.	Literature
4. Identify the target audience.	Survey results
5. Re-examine the social media goals set.	Knowledge gained during this study

Guidelines	Developed according to
6. Test various types of content on different social media platforms.	Knowledge gained during this study
7. Identify the platform from which social media data will be collected.	Literature
8. Make use of techniques or methods to gather social media data.	Framework created by AMEC (2013)
9. Analyze the social media data.	Knowledge gained during this study, as well as framework created by Jeffrey (2013) and Paine (2015)
10. Analyze the results and determine social return-on-investment.	Framework created by Jeffrey (2013) and Paine (2015)
11. Evaluate the process on a regular basis.	Framework created by Jeffrey (2013) and Paine (2015)

- A contribution to research being done in the particular area of social media.

Research being done on the use of social media, specifically when companies are involved, is still very new. This study contributed not only to research being done on social media, but also research considering the use of social media within companies and how companies can benefit by using social media platforms.

1. Align the company's social media goals with the company's goals/business objectives.

> *Avoid imitating social media platforms of competitive companies or companies in the same industry. Each company, employee, customer, product, service, etc. is unique.*
>
> *Companies should maximize efforts to benefit from the fact that promotion of posts on social media platforms are mostly free.*

2. Prioritize social media goals.

> *Start with smaller social media goals and try to achieve these goals before trying to achieve social media goals with a large scope.*

3. Identify a timeframe.

> *For each social media goal set, identify a timeframe in which the goal's progress must be measured and achieved by. For example, if a company wants their Facebook page to reach 5000 likes and they are currently on 600 likes set a timeframe of 6 months and try to get 1000 likes each month.*

4. Identify the target audience.

> *Identify the awareness of customers or clients regarding the company's social media presence. The company can make use of surveys to determine their target audience and to create a profile of their customers or clients.*

5. Re-examine the social media goals set.

> *Re-examine the social media goals set by the company to ensure that the social media goals can be achieved with the help of the identified target audience. Certain social media goals can only be achieved if content is generated by customers or clients.*
>
> *Companies should use Facebook and Twitter for marketing as these proved to be most effective.*

6. Test various types of content on different social media platforms.

> *By testing different types of content, for example text, images, video, etc. a company can identify which content yields the most user generated content from customers or clients. A good idea is to test which content is suitable for each different social media platform that the company is active on.*
>
> *Companies can also make use of bulk SMS campaigns, Yammer and WhatsApp messenger to engage with customers.*

7. Identify the platform from which social media data will be collected.

> *Identify the social media platform on which customers or clients are most active. Social media platforms offer dashboards which indicate the daily activity, not only of the company but also the activity of customers or clients.*

8. Make use of techniques or methods to gather social media data.

> *The identified social media platform from which data will be gathered might include techniques or methods that can be used to gather the user generated content. Social media platforms today offer API's that can be used to gather social media data.*

9. Analyze the social media data.

> *Depending on the information that the company wants to gain from the social media data different text- and data-mining techniques can be used to analyze the social media data. Existing software allows companies to do text- and data-mining to determine sentiment, opinions, etc.*

10. Analyze the results and determine social return-on investment.

> *Results should be compared to previous results obtained from a different timeframe. Social return-on-investment should also be measured by calculating the cumulative time benefits and costs.*

11. Evaluate the process on a regular basis.

> *By evaluating the process on a regular basis and examining new techniques and methods for gathering and analyzing social media data customer relationship management can adapt new and improved strategies.*
>
> *Companies can create opportunities in areas where customers/clients do not have access to the Internet or technology. For example, place a touch screen computer in a shop which is located in a rural area where customers or clients can then experience social media and give feedback regarding products or services.*

Figure 7.2 - Basic framework that can be used by companies.

The answer to the research question of this study, as well as secondary objectives that have been set are discussed next.

7.5 Answering the research question and achieving secondary objectives

The research question, which should be answered in this study, is: Which techniques can be used by a company to ensure that social media is utilized effectively?

Each company is unique, not only the type of company, but the company's customers or clients as well. When choosing a technique to ensure that social media is utilized effectively, the techniques should be chosen in such a manner that they are befitting to what the company wants to achieve by using social media. If a company wants to measure the activity and awareness of customers/clients on the company's social media platforms, techniques that can be used include reporting tools, advertisement managers, Application Programming Interfaces (API's) of different social media platforms, social media management systems (for example Hootsuite) and built-in dashboards offered by social media platforms.

A company can also make use of text-, data- and web-mining techniques to analyse data gathered from social media platforms and to perform sentiment analysis and opinion mining. Sentiment analysis and opinion mining can contribute to understanding a company's customer or clients better. This can ensure that different strategies are applied when posting content on social media platforms.

To answer the research question the researcher had to complete secondary objectives (SO). All the secondary objectives have been achieved by the researcher. The secondary objectives listed in Chapter 1 (§1.4) follow, as well as how these secondary objectives were met.

SO1: Review literature on the use of social media in companies.

Social media was defined and the origin of social media was discussed. Different public external social media platforms, including Facebook, Twitter, LinkedIn, YouTube and Google+ were investigated. A comparison between the different social media platforms was carried out and general problems and limitations gathered from literature were discussed. The use of social media to build customer relationships was also investigated. Social media measurement processes currently available and being used by companies were also investigated. This study objective was achieved as indicated in Chapter 2.

SO2: Research different data- and text-mining techniques that can be used to analyse social media data and to gather social media data.

Data- and text mining in the business environment was investigated. Sentiment analysis and opinion mining can be done on social media data. Different methods and techniques can be used to mine text and various terminologies, such as text and web mining are often associated with data mining. Data-mining techniques, which could be used for sentiment analysis and opinion mining, were identified. For this study, classification models were used. This study objective was achieved and gave rise to the literature chapter, Chapter 3.

SO3: Gather and collect non-structured data, for instance "tweets" of customers from a certain company or within a certain business area.

Tweets from three different South African companies were collected over a period of a month. Different techniques were identified, which could be used to easily gather tweets. For this study, Twitter's API was used in conjunction with Google sheets. The process used to gather tweets was explained in Chapter 4, section 4.4.2.

SO4: Determine patterns, such as identification of topic keywords, which can be found on a company's social media sites, as well as user-generated content.

Different classification models were used to determine patterns and keywords from tweets gathered. The larger the data set, the more information can be gathered after analyses have been done. For example, 620 tweets were gathered concerning Company Y, a decision tree was used and keywords such as great, honour, win, fun, etc. could be associated with this company. By using a software platform, such as Rapidminer Studio, determining patterns and keywords from user-generated content is easy and fast depending on the built process being used. This was discussed in Chapter 6.

SO5: Use data- and text-mining techniques and methods (identified in S02) to determine sentiment and extract hidden information.

Data mining techniques identified by means of SO2 were used to determine patterns and keywords from tweets gathered (SO3). Classification models were also used in this secondary objective to achieve results regarding sentiment and opinions of customer or clients. Rapidminer Studio and the AYLIEN Text Analysis extension were used as a tool to apply data- and text-mining techniques to tweets gathered. The results obtained after analysing tweets for sentiment and opinions are presented in Chapter 6.

SO6: Create a connection between text sentiment and public opinion. This will be done by determining if a user's message articulates a negative or positive opinion regarding a certain company, brand, product or service.

A clear distinction could be made between a negative or positive tweet. Tweets were also analysed according to the opinion of a customer or client. Opinions were classified as subjective, a reflection of the customer's opinion, or objective, an expression of fact. After using the AYLIEN Text Analysis extension and classifying a tweet as positive or negative, the tweet was then placed through a process where one of three classification models was used to predict whether the sentiment of the tweet was done correctly. This secondary objective was achieved successfully and is elaborated on in Chapter 6.

SO7: Determine how social media is used in companies today by developing a questionnaire.

After the literature was reviewed (SO2 and SO3), a questionnaire was developed and sent to different departments of different companies. The questionnaire was pre-tested and approved by Statistical Consultation Services, North-West University, Potchefstroom campus. A total of 122 questionnaires were completed. The process followed during the development of the questionnaire is outlined in Chapter 4. Results regarding how social media is currently used in companies are presented in Chapter 5.

SO8: Create a basic social media framework to facilitate the analyses of social media data that can be gathered from company's social media platforms. The framework is not a standard that should be followed, but rather a recommended approach and tool.

From the literature and the results obtained from the survey, a basic framework was compiled, which can be used as a guideline for companies starting to use social media. The framework is depicted in Figure 7.2 on page 166.

SO9: Draw a conclusion on the effectiveness of utilizing social media in a company.

A conclusion was drawn regarding the effectiveness of utilizing social media in a company in Chapter 7 of this dissertation. An elaboration of the findings of this study is also given in this chapter.

The limitations documented for this study are subsequently discussed.

7.6   Limitations of this study

The limitations of this study are as follows:

- Even though the questionnaires were distributed to a large number of companies, only 122 questionnaires were received.
- The quality of information obtained from open-ended questions could have been improved if the researcher could follow-up open-ended questions that have not been answered.
- The techniques, strategies and framework investigated have not been tested in practice.

- Other social media data, for example Facebook user-generated content, can also be used to determine sentiment and opinions of customers. The reason why Twitter data was used for this study is that Twitter data can be collected for free. Companies can buy data gathered from Facebook, but new techniques for gathering Facebook data are currently being developed and should be considered.

From this study, a few ideas regarding future work emerged and are discussed next.

7.7   Future work and research

Future work and research, which can be considered include:

- Processes used to identify different personalities that use social media platforms can be identified and elaborated. This can be useful as a company would be able to determine which types of personality are visiting the company's social media platforms and how social media content can be developed according to these personalities.

- Problems and challenges that occur when a company uses social media can be addressed and solutions can be developed for these problems and challenges. From this study results indicated that social media is useful for a company; a company should not be prevented by problems and challenges from using social media. For example, social media fatigue is becoming a popular challenge for companies and from available literature, the problem has not yet been addressed.

- Results obtained by means of surveys indicated that the size of companies should be considered when using social media platforms. Companies of different sizes do not always have the same resources and time, which can be used to improve the company's social media presence. This can be investigated in more depth to determine which resources are of absolute importance to a company, no matter the size, when using social media platforms.

- The basic framework developed in this study can be elaborated on, tested and adjusted according to the type of company and/or type of customer.

- Results obtained by means of surveys indicated that companies are still in the process of developing techniques to fully measure the social return on investment and the contribution that social media is making to the growth of a company's customer base, as well as the growth of sales.

- Graphical social media data posted on, for example Facebook and Instagram can be analysed and image-processing techniques, which determine the emotion of the customer, type of product, brand name, etc. can be investigated.

- If possible, similar research should be conducted on a regular basis because it can contribute to improving the knowledge regarding the use of social media in companies as this type of research is still expanding and innovative.

7.8   Conclusion

In this chapter the aim was to present the findings gained from the survey and the analysed Twitter data. The research question and secondary objectives set for this study have been answered. The general information regarding companies who participated in this study was summarized. The effectiveness of different social media platforms was also discussed and summarized. Social media strategies and metrics used for gathering analytics and data from social media were discussed and a basic framework was developed, which can be used as a guideline or tool by companies.

Social media data play a significant role in the framework and sentiment analysis, and opinion mining can be performed on this data to gain insight into a company's customers or clients. Results obtained from the analysis of Twitter data were discussed and companies are encouraged to use software platforms, such as Rapidminer Studio to determine sentiment and opinions. The processes built in this software platform for this study can be adapted for future research. A limitation of this study was that the techniques, methods, strategies and developed framework could not be tested in practice.

This study contributed to improving the knowledge regarding the value of social media and enriched the research currently being done in the field of social media. This study can serve as a basis for future work and research in social media and particularly the use of social media in companies. This study confirms that social media can be utilized effectively and to the benefit of companies, depending on the techniques, strategies and frameworks that a company chooses to apply.

BIBLIOGRAPHY

ADEDOYIN-OLOWE, M., GABER, M. M. & STAHL, F. 2013. *Survey of Data Mining Techniques for Social Media Analysis* [Online]. Available: http://arxiv.org/vc/arxiv/papers/1312/1312.4617v1.pdf [Accessed 9 November 2015.

ADLER, P. S. & KWON, S. 2002. Social Capital: Prospects for a New Concept. *Academy of Management Review,* 29**,** 17-40.

AGRAWAL, R. & BATRA, M. 2013. A Detailed Study on Text Mining Techniques. *International Journal of Soft Computing and Engineering,* 2**,** 118 - 121.

AGRAWAL, R., IMIELINSKI, T. & SWAMI, A. Mining Association Rules between Sets of Items in Large Databases. 1993 ACM SIGMOD international conference on Management of data, 1993 New York, USA. 207-216.

AHMED, W. 2015. *Using Twitter as a data source: An overview of current social media research tools* [Online]. Available: http://blogs.lse.ac.uk/impactofsocialsciences/2015/07/10/social-media-research-tools-overview/ [Accessed 26 January 2016.

AL-AZMI, A. R. 2013. Data, Text, and Web Mining for Business Intelligence: A Survey. *International Journal of Data Mining & Knowledge Management Process,* 3**,** 1-21.

AL-JANABI, S. 2013. *Text Mining and Data Mining?* [Online]. Available: http://www.researchgate.net/post/Text_mining_and_Data_mining [Accessed 27 October 2015.

AMEC. 2013. *Unlocking Business Performance* [Online]. Available: http://amecorg.com/wp-content/uploads/2013/06/Social-Media-Valid-Framework2013.pdf [Accessed 19 January 2016.

AMEC. 2014. *AMEC's Social Media Measurement Framework: User Guide* [Online]. Available: http://www.social-media-measurement-framework.org/wp-content/uploads/2014/06/Social-Media-Measurement-Framework.pdf [Accessed 19 January 2016.

BALASUBRAMANIAM, N. 2009. *User-Generated Content* [Online]. Available: http://www.im.ethz.ch/education/FS09/iotsem09_proceedings.pdf#page=28 [Accessed 29 January 2015.

BALNAVES, M. & CAPUTI, P. 2010. *Introduction to quantitative research,* London, Sage Publications.

BEAL, V. 2016. *API Application Programming Interface* [Online]. Available:
http://www.webopedia.com/TERM/A/API.html [Accessed 22 September 2016.

BERRY, M. J. A. & LINOFF, G. S. 2004. *Data Mining Techniques For Marketing, Sales, and
Customer Relationship Management,* Indianapolis, Indiana, Wiley Publishing, Inc.

BERSON, A., SMITH, S. & THEARLING, K. 2005. *An Overview of Data Mining Techniques*
[Online]. Available: http://www.thearling.com/text/dmtechniques/dmtechniques.htm [Accessed
15 September 2015.

BRADSHAW, T. 2011. *Shame and share alike: Social media vie for attention* [Online].
Available: http://www.ft.com/cms/s/2/1671609a-d9f3-11e0-b199-
00144feabdc0.html#axzz3hkUJnE9k [Accessed 22 July 2015.

BRIGHT, L. F., KLEISER, S. B. & GRAU, S. L. 2014. Too much Facebook? An exploratory
examination of social media fatigue. *Computers in Human Behavior,* 44**,** 148-155.

BROOKE, C. 2002. What does it mean to be 'critical' in IS research? *Journal of Information
Technology,* 17**,** 49-57.

BROWN, M. 2012. *Data mining techniques* [Online]. Available:
http://www.ibm.com/developerworks/library/ba-data-mining-techniques/ba-data-mining-
techniques-pdf.pdf [Accessed 8 September 2015.

BUSINESSDICTIONARY. 2015. *Definition of customer relationship management* [Online].
Available: http://www.businessdictionary.com/definition/customer-relationship-management-
CRM.html [Accessed 6 February 2015.

CLEVER, N., KIRCHNER, A., SCHRAY, D. & SCHULTE, M. 2009. *User-generated content*
[Online]. Available: http://453.stilled.net/wp-content/uploads/2010/06/Eessay-user-generated-
content.pdf [Accessed 23 January 2015.

COLEMAN, J. S. 1988. Social Capital in the Creation of Human Capital. *The American Journal
of Sociology,* 94**,** 95-120.

CONSTANTINIDES, E. & FOUNTAIN, S. J. 2008. Web 2.0: Conceptual foundations and
marketing issues. *Journal of Direct, Data and Digital Marketing Practice,* 9**,** 231-244.

CORMODE, G. & KRISHNAMURTHY, B. 2008. *Key differences between Web 1.0 and Web 2.0*
[Online]. Available: http://firstmonday.org/article/view/2125/1972 [Accessed 2 June 2015.

CRESWELL, J. W. 2014. *Research Design - Qualitative, Quantitative, and Mixed Method Approaches,* London, SAGE Publications.

CREWS, T. B. & STITT-GOHDES, W. L. 2012. Incorporating Facebook and Twitter in a Service-Learning Project in a Business Communication Course. *Business Communication Quarterly,* 20**,** 1-4.

CVIJIKJ, I. P. & MICHAHELLES, F. 2011. *Understanding Social Media Marketing: A Case Study on Topics, Categories and Sentiment on a Facebook Brand Page* [Online]. Available: http://www.im.ethz.ch/publications/Pletikosa-MindTrek11 [Accessed 29 January 2015.

DALCHER, D. & BRODIE, L. 2007. *Successful IT Projects,* London, Thomson Learning.

DAUGHERTY, T., EASTIN, M. S. & BRIGHT, L. 2008. Exploring Consumer Motivations for Creating User-Genrated Content. *Journal of Interactive Advertising,* 8**,** 16-25.

DE VAULT, G. 2016. *Quantitative Research - Advantages and Disadvantages* [Online]. Available: https://www.thebalance.com/quantitative-research-advantages-and-disadvantages-2296728 [Accessed 22 September 2016.

DI MASSA, M. 2012. *Web 2.0 tools* [Online]. Available: http://www.iet.unipi.it/m.cimino/publications/2012,%20SEM,%20Web%202.0%20tools.pdf [Accessed 26 May 2015.

DICKINSON, B. & HU, W. 2015. Sentiment Analysis of Investor Opinions on Twitter. *Social Networking,* 4**,** 62-71.

DIJKMANS, C., KERKHOF, P. & BEUKEBOOM, C. J. 2015. A stage to engage: Social media use and corporate reputation. *Tourism Management,* 47**,** 58-67.

ELLISON, N. B., STEINFIELD, C. & LAMPE, C. 2007. The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication,* 12**,** 1143-1168.

ERDOĞMUŞ, I. E. & ĆIĆEK, M. 2012. The impact of social media marketing on brand loyalty. *Procedia – Social and Behavioral Sciences,* 58**,** 1353-1360.

FAN, W., WALLACE, L., RICH, S. & ZHANG, Z. 2006. Tapping the Power of Text Mining. *Communications of the ACM,* 49**,** 77-82.

GEHO, P. R., SMITH, S. & LEWIS, S. D. 2010. *Is Twitter a Viable Commercial Use Platform for Small Businesses? An Empirical Study Targeting Two Audiences in the Small Business Community* [Online]. Available: http://www.freepatentsonline.com/article/Entrepreneurial-Executive/243043184.html [Accessed 13 July 2015.

GILL, P., ARLITT, M., LI, Z. & MAHANTI, A. YouTube Traffic Characterization: A View From the Edge. 7th ACM SIGCOMM Conference on Internet Measurement, 2007 New York, USA. 15-28.

GLIEM, J. A. & GLIEM, R. R. 2003. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *2003 Midwest Research to Practice Conference in Adult, Continuing and Community Education.* Columbus, Ohio.

GOLDBERG, L. R. & VELICER, W. F. 2006. *Differentiating normal and abnormal personality,* New York, Springer.

GOLFARELLI, M., RIZZI, S. & CELLA, I. 2004. *Beyond Data Warehousing: What's Next in Business Intelligence?* [Online]. Available: http://www.cci.drexel.edu/faculty/song/dolap/dolap04/paper/p1-rizzi.pdf [Accessed 14 September 2015.

GONZALEZ, R., CUEVAS, R., MOTAMEDI, R. & REJAIE, R. 2012. *Google+ or Google-? Dissecting the Evolution of the New OSN in its First Year* [Online]. Available: http://arxiv.org/pdf/1205.5662.pdf [Accessed 10 July 2015.

GOODFELLOW, G. 2012. *Facebook in the Workplace: Findings of a Content Analysis of Canadian Arbitration Cases.* Master of Business Administration, University of Prince Edward Island.

GOOGLE. 2015. *About Google+ pages* [Online]. Available: https://support.google.com/plus/answer/1710600?hl=en [Accessed 13 July 2015.

GREENBERG, P. 2010. The impact of CRM 2.0 on customer insight. *Journal of Business & Industrial Marketing,* 25**,** 410-419.

GUARINO, N. 1995. Formal ontology, conceptual analysis and knowledge representation. *International Journal Human-Computer Studies,* 43**,** 625-640.

GUBA, E. G. & LINCOLN, Y. S. 1994. *Handbook of Qualitative Research,* Thousand Oaks, Sage Publications.

GUNDECHA, P. & LIU, H. 2012. *Mining Social Media: A Brief Introduction* [Online]. Available: http://www.public.asu.edu/~pgundech/book_chapter/smm.pdf [Accessed 10 July 2015.

HALLEBONE, E. & PRIEST, J. 2009. *Business & Management Research: Paradigms & Practices* [Online]. Available: http://www.nasirmehmood.com/articals/43.1332820806.Business%20and%20Management%20 Research%20Paradigms%20&%20Practices.pdf [Accessed 28 October 2015.

HAN, J., KAMBER, M. & PEI, J. 2012. *Data Mining: Concepts and Techniques,* Waltham, USA, Elsevier, Inc. .

HANNA, R., ROHM, A. & CRITTENDEN, V. L. 2011. We're all connected: The power of the social media ecosystem. *Business Horizons,* 54**,** 265-273.

HARIDAKIS, P. & HANSON, G. 2009. Social Interaction and Co-Viewing With YouTube: Blending Mass Communication Reception and Social Connection. *Journal of Broadcasting and Electronic Media,* 53**,** 317-335.

HARVEY, D. 1990. *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change,* Cambridge, Massachusetts, Blackwell Publishers.

HE, W., ZHA, S. & LI, L. 2013. Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management,* 33**,** 464-472.

HEALE, R. & TWYCROSS, A. 2015. Validity and reliability in quantitative studies. *Evid Based Nurs,* 18**,** 66-67.

HEGLAND, M. 2001. *Data mining techniques* [Online]. Available: http://www.researchgate.net/publication/231906788_Data_mining_techniques [Accessed 17 July 2015.

HESSE, E. 1980. *Revolutions and reconstructions in the philosophy of science,* Bloomington Indiana, University Press.

HILL, J. 2015. *6 Social Media Studies You Should Know About In 2016* [Online]. Available: https://us.searchlaboratory.com/2015/06/6-social-media-studies-you-should-know-about-in-2015/ [Accessed 26 January 2016.

HINES, K. n.d. *Google+ - How to Use Google+ for Business and Professional Branding* [Online]. Available: http://www.wordtracker.com/attachments/google-plus-final.pdf [Accessed 10 July 2015.

HO, R. 2006. *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS,* United States of America, Chapman & Hall/CRC.

HOTHO, A., NÙRNBERGER, A. & PAAB, G. 2005. *A Brief Survey of Text Mining* [Online]. Available:
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.210.158&rep=rep1&type=pdf [Accessed 9 July 2015.

HOWCROFT, D. & TRAUTH, E. M. 2004. *The Choice of Critical Information Systems Research* [Online]. Available: http://link.springer.com/chapter/10.1007%2F1-4020-8095-6_12#page-1 [Accessed 13 April 2015.

JACOBS, A. 2009. The Pathologies of Big Data. *Communications of the ACM,* 52**,** 36-44.

JAYACHANDRAN, S., SHARMA, S., KAUFMAN, P. & RAMAN, P. 2005. The Role of relational Information Processes and Technology Use in Customer Relationship Management. *Journal of Marketing,* 69**,** 177-192.

JEFFREY, A. 2013. *Social Media Measurement: A Step-by-Step Approach* [Online]. Available: http://www.instituteforpr.org/wp-content/uploads/Social-Media-Measurement-Paper-Jeffrey-6-4-13.pdf [Accessed 19 January 2016.

JELONEK, D. The Evolution of Customer Relationship Management System 19th International Conference on Computers, 2015 Zakynthos Island, Greece. 29 - 33.

JOHNSON, P. R. & YANG, S. 2010. Uses and gratifications of Twitter: An examination of user motives and satisfaction of Twitter use. *Communication Research,* 37**,** 473.

JOINSON, A. N. 'Looking at', 'Looking up' or 'Keeping up with' People? Motives and Uses of Facebook. CHI 2008 - Online Social Networks, 2008 Florence, Italy. 1027-1036.

JONES, S., MURPHY, F., EDWARDS, M. & JAMES, J. 2008. Doing things differently: advantages and disadvantages of web questionnaires. *Nurse Researcher,* 15**,** 15-26.

JOUBISH, M. F., KKURRAM, M. A., AHEMED, A., FATIMA, S. T. & HAIDER, K. 2011. Paradigms and Characteristics of a Good Qualitative Research. *World Applied Sciences Journal,* 12**,** 2082-2087.

JUSSILA, J. J., KÄRKKÄINEN, H. & LEINO, M. 2011. *Benefits of Social Media in Business-to-Business Customer Interface in Innovation* [Online]. Available:

https://dspace.cc.tut.fi/dpub/bitstream/handle/123456789/22017/jussila_benefits_of_social_media_in_business_to_business.pdf?sequence=1 [Accessed 5 February 2015.

KAIRAM, S., BRZOZOWSKI, M. J., HUFFAKER, D. & CHI, E. H. 2012. *Talking in Circles: Selective Sharing in Google+* [Online]. Available: http://www.researchgate.net/profile/David_Huffaker/publication/239761331_Talking_in_circles_selective_sharing_in_google/links/0deec536abc21ca266000000.pdf [Accessed 6 July 2015.

KANTARDZIC, M. 2011. *Data Mining: Concepts, Models, Methods and Algorithms,* Hoboken, New Jersey, John Wiley & Sons, Inc.

KAPLAN, A. M. & HAENLEIN, M. 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons,* 53**,** 59-68.

KÄRKKÄINEN, H., JUSSILA, J. & VÄISÄNEN, J. 2010. *Social media use and potential in business-to-business companies' innovation* [Online]. Available: http://dspace.cc.tut.fi/dpub/bitstream/handle/123456789/22019/karkkainen_social_media_use_and_potential.pdf?sequence=1 [Accessed 29 January 2015.

KENT, M. L. & TAYLOR, M. 1998. Building Dialogic Relationships Through the World Wide Web. *Public Relations Review,* 24**,** 321-334.

KIETZMANN, J. H., HERMKENS, K., MCCARTHY, I. P. & SILVESTRE, B. S. 2011. Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons,* 54**,** 241-251.

KLEIN, H. K. & MYERS, M. D. 2001. *Qualitative research in IS: issues and trends,* United Sates of America, Idea Group Publishing.

KRAUSS, S. E. 2005. Research Paradigms and Meaning Making: A Primer. *The Qualitative Report,* 10**,** 758-770.

KROEZE, J. H., MATTHEE, M. C. & BOTHMA, T. J. D. 2004. Differentiating between data-mining and text-mining terminology. *South-African Journal of Information Management,* 6**,** 1-14.

KRUMM, J., DAVIES, N. & NARAYANASWAMI, C. 2008. *User-Generated Content* [Online]. Available: http://www.computer.org/csdl/mags/pc/2008/04/mpc2008040010.pdf [Accessed 18 September 2015.

KUSHMERICK, N., WELD, D. S. & DOORENBOS, R. 1997. *Wrapper Induction for Information Extraction* [Online]. Available: http://www.isi.edu/info-agents/courses/iiweb/kushmerick-ijcai97.pdf [Accessed 22 October 2015.

KWAK, H., LEE, C., PARK, H. & MOON, S. What is Twitter, a Social Network or a New Media? 19th International conference on World Wide Web, 2010 San Jose, California. 435-444.

KWOK, L. & YU, B. 2013. Spreading Social Media Messages on Facebook: An Analysis of Restaurant Business-to-Consumer Communications. *Cornell Hospitality Quarterly,* 54**,** 84-94.

LAMPE, C., ELLISON, N. & STEINFIELD, C. A Familiar Face(book): Profile Elements as Signals in an Online Social Network. CHI 2007 2007 San Jose, California. 435-444.

LARDI, K. & FUCHS, R. 2013. *Social Media Strategy: A Step-by-Step Guide to Building Your Social Business,* Zürich, vdf Hochschulverlag AG

LEE, A. S. 1991. Integrating Positivist and Interpretive Approaches to Organizational Research. *Organization Science,* 2**,** 342 - 365.

LEGGAT, H. 2010. *Rebuild brand loyalty with social media* [Online]. Available: http://www.bizreport.com/2010/08/price-sensitive-shoppers-still-seeking-out-deals.html [Accessed 16 February 2015.

LEHTIMÀKI, T., SALO, J., HILTULA, H. & LANKINEN, M. 2009. *Harnessing web 2.0 for business to business marketing – Literature review and an empirical prospective from Finland* [Online]. Available: http://herkules.oulu.fi/isbn9789514291203/isbn9789514291203.pdf [Accessed 4 February 2015.

MACKENZIE, N. & KNIPE, S. 2006. *Research dilemmas: Paradigms, methods and methodology* [Online]. Available: http://www.iier.org.au/iier16/mackenzie.html [Accessed 21 March 2015.

MADRIGAL, D. & MCCLAIN, B. 2012. *Strengths and Weaknesses of Quantitative and Qualitative Research* [Online]. Available: http://www.uxmatters.com/mt/archives/2012/09/strengths-and-weaknesses-of-quantitative-and-qualitative-research.php [Accessed 22 September 2016.

MAGNO, G., COMARELA, G., SAEZ-TRUMPER, D., CHA, M. & ALMEIDA, V. 2012. *New Kid on the Block: Exploring the Google+ Social Graph* [Online]. Available: http://www.researchgate.net/profile/Diego_Saez-

Trumper/publication/263005450_New_Kid_on_the_Block_Exploring_the_Google_Social_Graph /links/0c9605398491ba9b60000000.pdf [Accessed 10 July 2015.

MANGOLD, W. G. & FAULDS, D. J. 2009. Social media: the new hybrid element of the promotion mix. *Business Horizons,* 52**,** 357-365.

MCCARRA, D. 2011. *Google bans promotions and competitions on Google+ Pages for brands* [Online]. Available: http://sociable.co/social-media/google-bans-promotions-and-competitions-on-google-pages-for-brands/ [Accessed 10 July 2015.

MCGREGOR, S. L. T. & MURNANE, J. A. 2010. Paradigm, methodology and method: intellectual integrity in consumer scholarship. *International Journal of Consumer Studies,* 34**,** 419-427.

MERRIL, T., LATHAM, K., SANTALESA, R. & NAVETTA, D. 2011. *Social Media: The Business Benefits May Be Enormous, But Can the Risks – Reputational, Legal, Operational – Be Mitigated?* [Online]. Available: http://www.acegroup.com/us-en/assets/ace-progress-report-social-media.pdf [Accessed 26 January 2016.

MEYERS, L. S., GAMST, G. & GUARINO, A. J. 2009. *Data Analysis Using SAS Enterprise Guide,* United States of America, New York, Cambridge University Press.

MONCRIEF, W. C., MARSHALL, G. W. & RUDD, J. M. 2015. Social media and related technology: Drivers of change in managing contemporary sales force. *Business Horizons,* 58**,** 45-55.

MOSS, L. T. & ATRE, S. 2003. *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications,* Boston, USA, Pearson Education, Inc. .

MUKHERJI, P. & ALBON, D. 2010. *Research Methods in Early Childhood – An Introductory Guide,* Great Britain, Sage Publications.

MURUGESAN, S. 2007. Understanding Web 2.0. *IT Pro***,** 34-41.

MYERS, M. D. 1997. *Qualitative Research in Information Systems* [Online]. Available: http://www.jstor.org.nwulib.nwu.ac.za/stable/249422?origin=crossref&seq=1#page_scan_tab_contents [Accessed 17 March 2015.

NANDI, G. & DAS, A. 2013. A Survey on Using Data Mining Techniques for Online Social Network Analysis. *International Journal on Computer Science Issues,* 56**,** 162-167.

NEELAMEGAM, S. & RAMARAJ, E. 2013. Classification algorithm in Data mining: An Overview. *International Journal of P2P Network Trends and Technology (IJPTT),* 4**,** 369-374.

NOVAS, C. 2015. *Social Networking Sites: Benefits, Problems and "Facebook Depression"* [Online]. Available: http://center4research.org/child-teen-health/1-general-health-and-mental-health/social-networking-sites-benefits-problems-and-facebook-depression/ [Accessed 13 July 2015.

OATES, B. J. 2006. *Researching Information Systems and Computing,* United Kingdom, Sage Publications.

ORLIKOWSKI, W. J. & BAROUDI, J. J. 1991. Studying Information Technology in Organizations: Research Approaches and Assumptions. *Information Systems Research,* 2**,** 1-28.

PAINE, K. D. 2015. *Theory Meets Reality: Implementing PR & Social Media Measurement Standards.* [Online]. Available: http://painepublishing.com/wp-content/uploads/2015/04/NJ-Social-Media-Summit.pdf [Accessed 19 January 2016.

PAK, A. & PAROUBEK, P. 2010. *Twitter as a Corpus for Sentiment Analysis and Opinion Mining.* [Online]. Available: http://lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf [Accessed 22 July 2015.

PANG, B. & LEE, L. 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* 2**,** 1–135.

PANIAGUA, J. & SAPENA, J. 2014. Business performance and social media: Love or hate? *Business Horizons,* 57**,** 719-728.

PAOLILLO, J. C. 2008. *Structure and Network in the YouTube Core* [Online]. Available: http://www.computer.org/csdl/proceedings/hicss/2008/3075/00/30750156.pdf [Accessed 10 July 2015.

PARVATIYAR, A. & SHETH, J. N. 2001. Customer Relationship Management: Emerging Practice, Process, and Discipline. *Journal of Economic and Social Research,* 3**,** 1-34.

PETRE, R. 2013. Data Mining Solutions for the Business Environment. *Database Systems Journal,* 4**,** 21-29.

PICARDO, A. 2013. *What To Do About Social Media Fatigue* [Online]. Available:
http://blogs.gartner.com/richard-fouts/2013/04/24/what-to-do-about-social-media-fatigue/
[Accessed 27 January 2015.

RAPIDMINER 2014. Rapidminer Studio Manual. Rapidminer.

READER'S DIGEST OXFORD COMPLETE WORDFINDER 1993. Definition of the terms social
and media. 1 ed. London: The Reader's Digest Association Limited.

REINARTZ, W., KRAFFT, M. & HOYER, W. D. 2004. The Customer Relationship Management
Process: Its Measurement and Impact on Performance. *Journal of Marketing Research,* 41**,**
293-305.

RIDEOUT, H. 2014. *Employee Risk Management: How to protect your business reputation and
reduce legal liability,* Philadelphia, United States, Kogan Page Limited.

ROBERTS, P. & PRIEST, H. 2006. Reliability and validity in research. *Nursing Standard,* 20**,**
41-45.

ROHRA, G. & SHARMA, M. 2012. *Social CRM - Possibilities and Challenges* [Online].
Available:http://www.tcs.com/SiteCollectionDocuments/White%20Papers/ConnectedMarketing_
Whitepaper_Social_CRM%E2%80%93Possibilities_Challenges_0912-1.pdf [Accessed 10
February 2016.

ROSS, C., ORR, E. S., SISIC, M., ARSENEAULT, J. M., SIMMERING, M. G. & ORR, R. R.
2009. Personality and motivations associated with Facebook use. *Computers in Human
Behavior,* 25**,** 578-586.

RUSSELL, M. A. 2014. *Mining the Social Web,* Sebastopol, California, O'Reilly Media, Inc.

RYGIELSKI, C., WANG, J. & YEN, D. C. 2002. Data mining techniques for customer
relationship management. *Technology in Society,* 24**,** 483-502.

SAKAKI, T., OKAZAKI, M. & MATSUO, Y. 2010. Earthquake Shakes Twitter Users: Real-time
Event Detection by Social Sensors. *19th International conference on World Wide Web.* San
Jose, California.

SKEELS, M. M. & GRUDIN, J. 2009. *When Social Networks Cross Boundaries: A Case Study
of Workplace Use of Facebook and LinkedIn* [Online]. Available:
http://research.microsoft.com/en-
us/um/people/jgrudin/publications/newwave/socialnetworking2009.pdf [Accessed 4 June 2015.

SMYTH, H. J. & MORRIS, P. W. G. 2007. An epistemological evaluation of research into projects and their management: Methodological issues. *International Journal of Project Management,* 25**,** 423-436.

SOCIAL MEDIA TODAY & LEADER NETWORKS. 2013. *The Socially Enabled Enterprise* [Online]. Available: http://www.leadernetworks.com/2013/09/what-it-means-to-become-a-socially-enabled-enterprise-research-study.html [Accessed 6 July 2015.

SPIEGEL, B. 2014. *Big Data 2.0 – YouTube Meets Text Analytics – Part 1* [Online]. Available: http://www.clickz.com/clickz/column/2380528/big-data-20-youtube-meets-text-analytics-part-1 [Accessed 21 October 2015.

STAHL, B. C. 2008. The ethical nature of critical research in information systems. *Information Systems Journal* 18**,** 137 - 163.

STATISTICS SOUTH AFRICA 2012. *Standard Industrial Classification of all Economic Activities (SIC),* Pretoria, South Africa, Statistics South Africa.

STELZNER, M. 2014. *LinkedIn Publishing Platform: What Marketers Need to Know* [Online]. Available: http://www.socialmediaexaminer.com/linkedin-publishing-platform-with-stephanie-sammons/ [Accessed 10 July 2015.

SUDERSAN, P. 1998. *Habermas and Critical Social Theory* [Online]. Available: http://www.unipune.ac.in/snc/cssh/ipq/english/IPQ/21-25%20volumes/25%2002/PDF/25-2-7.pdf [Accessed 14 April 2015.

TABACHNICK, B. G. & FIDELL, L. S. 2013. *Using Multivariate Statistics,* United States of America, Pearson Education, Inc.

TANNER, J. & RAYMOND, M. A. 2010. *Principles of Marketing* [Online]. Available: http://www.web-books.com/eLibrary/NC/B0/B64/025MB64.html [Accessed 2 February 2015.

TAYLOR, J. K. & CIHON, C. 2004. *Statistical Techniques for Data Analysis,* Boca Raton, Florida, Chapman & Hall/CRC.

TECHNOPEDIA. 2011. *Social Media Fatigue* [Online]. Available: http://www.techopedia.com/definition/27372/social-media-fatigue [Accessed 27 January 2015.

THEW, D. 2008. LinkedIn – a user's perspective. *Business Information Review,* 25**,** 87-90.

TORR, D. 2014. *7 Social Media Challenges And Business Opportunities For Social Media Managers* [Online]. Available: http://blog.hootsuite.com/social-media-challenges-small-businesses-face/ [Accessed 20 January 2016.

TRAINOR, K. J. 2013. Relating Social Media Technologies to Performance: A Capabilities-Based Perspective. *Journal of Personal Selling & Sales Management,* 32**,** 317-331.

TRAINOR, K. J., ANDZULIS, J., RAPP, A. & AGNIHOTRI, R. 2014. Social media technology usage and customer relationship performance: A capabilities-based examination of social CRM. *Journal of Business Research,* 67**,** 1201-1208.

TRAVIS, J. 1999. *Exploring the Constructs of Evaluative Criteria for Interpretivist Research* [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.4.1994&rep=rep1&type=pdf [Accessed 24 March 2015.

TSIPTSIS, K. & CHORIANOPOULOS, A. 2009. *Data Mining Techniques in CRM: Inside Customer Segmentation,* West Sussex, United Kingdom, Wiley & Sons, Ltd.

VAN ZYL, G. 2015. *INFOGRAPHIC: SA social media by the numbers* [Online]. Available: http://www.fin24.com/Tech/Multimedia/INFOGRAPHIC-SA-social-media-by-the-numbers-20150917 [Accessed 27 June 2016.

WANG, R. 2012. *Monday's Musing: avoiding Social Media Fatigue Through Engagement* [Online]. Available: http://www.forbes.com/sites/raywang/2012/05/21/mondays-musing-avoiding-social-media-fatigue-through-engagement/ [Accessed 5 February 2015.

WATERS, R. D., BURNETT, E., LAMM, A. & LUCAS, A. 2009. Engaging stakeholders through social networking: How nonprofit organizations are using Facebook. *Public Relations Review,* 35**,** 102-106.

WATERS, R. D. & JONES, P. M. 2011. Using Video to Build an Organization's Identity and Brand: A Content Analysis of Nonprofit Organizations' YouTube Videos. *. Journal of Nonprofit & Public Sector Marketing,* 23**,** 248-268.

WE ARE SOCIAL. 2016. *Digital In 2016* [Online]. Available: http://wearesocial.com/uk/special-reports/digital-in-2016 [Accessed 25 January 2016.

WEAVER, K. & JOANNE, K. 2006. Understanding paradigms used for nursing research. *Journal of Advanced Nursing,* 53**,** 459-469.

WIEGAND, N. 2015. *Social Media Advantages and Disadvantages in Business* [Online]. Available: http://everydaylife.globalpost.com/social-media-advantages-disadvantages-business-21665.html [Accessed 21 October 2015.

WILLIS, J. W. 2007. *Foundations of Qualitative Research – Interpretive and Critical Approaches,* London, Sage Publications.

WORLD WIDE WORX. 2016. *Instagram, Facebook, make waves in SA* [Online]. Available: http://www.worldwideworx.com/socialsa2016/ [Accessed 27 June 2016.

WU, X., KUMAR, V., QUINLAN, J. R., GHOSH, J. & YANG, Q. 2008. Top 10 algorithms in data-mining. *Knowledge Information Systems,* 14**,** 1-37.

YE, N. 2003. *The Handbook of Data-mining,* London, Lawrence Erlbaum Associates Publishers.

ZHANG, M., JANSEN, B. J. & CHOWDHURY, A. 2011. Business engagement on Twitter: a path analysis. *Electron Markets,* 21**,** 161-175.

ANNEXURE A - QUESTIONNAIRE

If your company is currently using social media, please take time to complete the following questionnaire.

Your participation in this questionnaire will contribute to research being done regarding the use of social media in companies. The study focuses on how social media can be used to improve customer relationship management, and the techniques used to extract sentiments and opinions from social media data.

Rest assured that the answers given by you will be kept anonymous at all times.

Please answer the following questions.

SECTION A - GENERAL INFORMATION
A_1 In which of the following business areas are you mainly employed in?

1	2	3	4	5	6
Retail, sales and marketing	Banking, finance and accounting	Insurance	Information technology and communications	Legal	Government

A_2 Roughly how many full-time employees are currently working for the company?

1	2	3	4
1 – 9 (micro enterprise)	10 – 49 (small enterprise)	50 – 249 (medium-sized enterprise)	250+ (large enterprise)

A_3 How useful do you think that social media has been for the company?

1	2	3	4
Not useful	Moderately useful	Useful	Very useful

SECTION B - SOCIAL MEDIA PLATFORMS
B_1 Are you allowed to visit external social media platforms, e.g. Facebook, Twitter, LinkedIn, etc. during working hours?

1	2
Yes	No

B_2 Does the company have social media training across the company to ensure that employees understand how to engage on social media platforms and consistently represent the company?

1	2
Yes	No

B_3 How long has the company been using social media platforms?

1	2	3	4	5	6
1 – 3 months	4 – 8 months	9 – 12 months	1 – 2 years	3 – 4 years	5+ years

B_4_1 Please rate the effectiveness of the following social media platforms used for marketing by the company?

		1	2	3	4	5
		Not effective	Moderately effective	Effective	Extremely effective	Not applicable
B_4_1_1	Facebook					
B_4_1_2	Twitter					
B_4_1_3	LinkedIn					
B_4_1_4	Google+					
B_4_1_5	YouTube					
B_4_1_6	Pinterest					
B_4_1_7	Instagram					
B_4_1_8	Tumblr					

B_4_2 Please rate the effectiveness of the following social media platforms used for branding by the company?

		1	2	3	4	5
		Not effective	Moderately effective	Effective	Extremely effective	Not applicable
B_4_2_1	Facebook					
B_4_2_2	Twitter					
B_4_2_3	LinkedIn					
B_4_2_4	Google+					
B_4_2_5	YouTube					
B_4_2_6	Pinterest					
B_4_2_7	Instagram					
B_4_2_8	Tumblr					

B_4_3 Please rate the effectiveness of the following social media platforms used for customer relationship management by the company?

		1	2	3	4	5
		Not effective	Moderately effective	Effective	Extremely effective	Not applicable
B_4_3_1	Facebook					
B_4_3_2	Twitter					
B_4_3_3	LinkedIn					
B_4_3_4	Google+					
B_4_3_5	YouTube					
B_4_3_6	Pinterest					
B_4_3_7	Instagram					
B_4_3_8	Tumblr					

B_4_4 Are there any other social media platforms used by the company that is not listed in the previous questions? (Text Response)

B_5_1 How does the company determine awareness of customers and clients regarding social media platforms being used? (Text Response)

B_5_2 According to the company's perspective, rate how aware customers and clients are regarding social media platforms being used.

1	2	3	4
Not aware	Moderately aware	Aware	Extremely aware

B_6_1 According to your opinion rate each of the following uses of social media by the company.

		1	2	3	4
		Almost never used	Moderately used	Regularly used	Almost always used
B_6_1_1	To advertise products/services.				
B_6_1_2	To offer promotional offers/items.				
B_6_1_3	To increase brand awareness.				
B_6_1_4	To gain feedback from customers.				
B_6_1_5	To engage (in conversation) with customers.				
B_6_1_6	For business to business purposes e.g. LinkedIn.				
B_6_1_7	To analyse the competition in the industry.				

B_6_2 List any other uses of social media platforms which is not listed above. (Text Response)

B_7 Indicate your level of agreement regarding the following statements.

		1	2	3	4
		Strongly disagree	Disagree	Agree	Strongly Agree
B_7_1	Social media is effective for brand awareness.				
B_7_2	Social media has revolutionised the marketing of a product/service.				
B_7_3	Social media is effective for selling a product/service.				
B_7_4	Social media is effective for advertising a product/service.				
B_7_5	Social media is great or engaging with customers.				

B_8 Indicate your level of agreement with the following statement: This social media platform allows a customer to express sentiment and his/her opinion.

		1	2	3	4	5
		Strongly disagree	Disagree	Agree	Strongly Agree	Not applicable
B_8_1	Facebook					
B_8_2	Twitter					
B_8_3	LinkedIn					
B_8_4	Google+					
B_8_5	YouTube					
B_8_6	Pinterest					
B_8_7	Instagram					
B_8_8	Tumblr					

B_9 Is your company paying to promote posts on social media platforms?

1	2
Yes	No

B_10_1 According to your opinion does different social media platforms attract a different audience?

1	2
Yes	No

B_10_2 How does the company determine the audience of their social media platforms? (Text Response)

B_11 What is a limitation that the company has experienced on a social media platform and how did the company overcome this? (Text Response)

B_12_1 Has the company hired or appointed employees specifically to manage the company's social media platforms?

1	2
Yes	No

B_12_2 If you have answered Yes to the question 12.1. have the employees appointed to manage the social media platforms received training or own a certain qualification, if Yes please specify. (Text Response)

B_12_3 Are these employee(s) appointed in-house or outsourced?

1	2
In-house	Outsourced

SECTION C - SOCIAL MEDIA STRATEGY

C_1 What are the goals that the company want to achieve by using social media platforms? (Text Response)

C_2 According to your opinion rate how integrated the company's social media goals are with the overall business objectives.

1	2	3	4
Not integrated	Moderately integrated	Integrated	Fully integrated

C_3 What are the steps that the company follows to encourage social media followers to become customers/clients? (Text Response)

C_4_1 Does the company make use of a framework or strategy when using social media platforms?

1	2
Yes	No

C_4_2_1 If you have answered Yes to question 4.1. please name the framework or strategy being used. (Text Response)

C_4_2_2 Rate the effectiveness of the named framework or strategy.

1	2	3	4	5
Not effective	Moderately effective	Effective	Very effective	Not applicable

C_5 Does your company make use of a dashboard to manage the different social media platforms?

1	2
Yes	No

C_6 How is your company monitoring competitive companies, products or services? (Text Response)

C_7 According to your opinion is there a lack of control on the information being dispersed about the company, when using a social media platform?

1	2
Yes	No

C_8_1 According to your opinion should the company be using social media more often in its day-to-day operations?

1	2
Yes	No

C_8_2 If you have answered Yes to question 8.1. please state why you as employee feel that the company should be using social media more in its day-to-day operations. (Text Response)

C_8_3 Please express your level of agreement regarding the following statements on why the company does not use social media platforms more often.

		1 Strongly disagree	2 Disagree	3 Agree	4 Strongly Agree	5 Not applicable
C_8_3_1	Not enough time.					
C_8_3_2	Not enough resources e.g. staff, finances, etc.					
C_8_3_3	Don't think that social media is useful to the organization.					
C_8_3_4	Have insufficient knowledge on how to use social media so the company will benefit.					

C_8_4 Please list any other reasons, than those listed above, why the company does not make use of social media platforms more often? (Text Response)

C_9 How does your company handle a social media crisis? For example if one of the social media platforms of the company is hacked. (Text Response)

SECTON D - SOCIAL MEDIA METRICS

D_1_1 Has the company established metrics to track social media efforts to the company's objectives and goals?

1	2
Yes	No

D_1_2 If you have answered Yes to question 1.1. what metrics are the company using to track social media efforts? (Text Response)

D_2_1 If the company has a website, is the company monitoring and gathering analytics and data from the website?

1	2
Yes	No

D_2_2 If you have answered Yes to question 2.1. how is the company gathering analytics and data from the website? (Text Response)

D_3_1 How often is content posted on social media platforms?

1	2	3	4	5
Multiple times per day	Daily	Weekly	Every fortnight	Monthly

D_3_2 Please rate the effectiveness of the type of content posted on social media platforms (where applicable).

		1	2	3	4	5
		Not effective	Moderately effective	Effective	Very effective	Not applicable
D_3_2_1	Text					
D_3_2_2	Images, photo's etc.					
D_3_2_3	Video					

D_4_1 According to your opinion do you agree that it is a good idea to link Twitter posts to automatically post to the company's Facebook page?

1	2	3	4
Strongly disagree	Disagree	Agree	Strongly agree

D_4_2 How does the company pull analytics and data from Twitter? (Text Response)

D_4_3 How does the company pull analytics and data from Facebook? (Text Response)

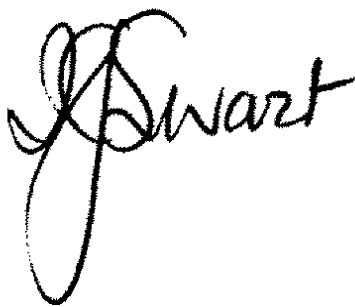**D_5** How does the company measure social return on investment (ROI)? (Text Response)

ANNEXURE B – LETTER CONFIRMING LANGUAGE EDITING

This serves to confirm that I, Isabella Johanna Swart, registered with and accredited as professional translator by the South African Translators' Institute, registration number 1001128, language edited the following dissertation:

Utilizing social media to the benefit of companies

by

Lizani Botha

Dr Isabel J Swart   Date: 17 November 2016

23 Poinsettia Close Van
der Stel Park Dormehlsdrift
GEORGE
6529
Tel: (044) 873 0111
Cell: 082 718 4210
e-mail: isaswart@telkomsa.net

ANNEXURE C – FIRST PAGE OF TURNITIN REPORT

Turnitin Originality Report

22206914:L_Botha_2016_11_17.docx by
LIZANI BOTHA

From Dissertation tests (fd3c1878-9efb-
4672-ba45-856c344b82d6)

Similarity Index	Similarity by Source	
16%	Internet Sources:	13%
	Publications:	5%
	Student Papers:	6%

Processed on 17-Nov-2016 11:12
SAST
ID: 738817326
Word Count: 67250

sources:

1   1% match (Internet from 26-May-2016)
http://dspace.nwu.ac.za/bitstream/handle/10394/10509/Nkone_MS.pdf?isAllowed=y&sequence=1

2   1% match (student papers from 06-Nov-2016)
Submitted to CTI Education Group on 2016-11-06

3   < 1% match (Internet from 23-Aug-2008)
http://www.functionalneurology.it/materiale_cic/25_XVIII_3/52_cinnarizine/index.html

4   < 1% match (Internet from 15-May-2014)
http://www.docstoc.com/docs/10961467/Data-Mining

5   < 1% match (Internet from 23-May-2014)
http://uzspace.uzulu.ac.za/bitstream/handle/10530/1237/THE%20ROLE%20OF%20SOCIAL%20MEDIA%20IN%20THE%20MARKETIN(sequence=1

6   < 1% match (Internet from 16-Dec-2014)
http://www.mis.boun.edu.tr/gulser/index_files/DM%20Techniques%20%26%20CRM_1.pdf

7   < 1% match (student papers from 25-Nov-2015)
Submitted to North West University on 2015-11-25

8   < 1% match (Internet from 09-Mar-2016)
http://www.celsi.sk/media/research_reports/celsi-rr1.pdf

9   < 1% match (Internet from 18-Jan-2015)
http://www.docstoc.com/docs/116854446/Data-Mining-Techniques-for-Marketing

10   < 1% match (Internet from 25-Feb-2014)
http://www.public.asu.edu/~pgundech/book_chapter/smm.pdf

11   < 1% match (Internet from 20-Mar-2014)
http://dspace.nwu.ac.za/bitstream/handle/10394/7004/Grey_J.pdf?sequence=2

12   < 1% match (Internet from 27-Jan-2014)
http://herkules.oulu.fi/isbn9789514291203/isbn9789514291203.pdf

13   < 1% match ()
http://www.i-pensieri.com/gregr/papers/DMHandbookPrediction.pdf

14   < 1% match (student papers from 20-May-2014)
Submitted to North West University on 2014-05-20

15   < 1% match (publications)
Social Media and Local Governments, 2016.

16   < 1% match (Internet from 13-Jun-2014)
http://iprrc.org/docs/IPRRC_16_Proceedings.pdf

17   < 1% match (student papers from 11-Nov-2015)
Submitted to North West University on 2015-11-11

18   < 1% match (Internet from 30-Nov-2015)
http://clok.uclan.ac.uk/10563/2/Ogunbiyi%20Oyedolapo%20Final%20e-Thesis%20%28Master%20Copy%29.pdf