Evaluating customer satisfaction using sentiment analysis:

A case study for mobile and fixed internet service providers in Zimbabwe

By

Martin Chamambo

(C19136264V)

A Research Dissertation Submitted in Partial Fulfilment of the

Requirements for the Degree of

Master of Science in Data Analytics

Graduate Business School

Chinhoyi University of Technology

Zimbabwe

Supervisor(s)

Mr H Mazhokota

April 2021

CHINHOYI UNIVERSITY OF TECHNOLOGY

Abstract

The focus of the study was to evaluate customer satisfaction using sentiment analysis in mobile and fixed internet service providers in Zimbabwe. Fixed internet and mobile Internet service providers have been critical in keeping the economy running under the lockdown measures effected by the government by providing business-critical connectivity and resilience, facilitating teleworking arrangements, e-banking, e-commerce and keeping individuals and societies connected and informed. As a result, many telecom players providing broadband have benefited from a surge in the traffic of data. This surge has caused a strain in customer service departments calling for new effective methods that can handle this surge and the real time nature of the support needed by customers. The main objective of the study was to measure customer satisfaction from tweets using sentiment analysis. A comprehensive review of both theoretical and empirical literature review related to this study was carried out. The study adopted the descriptive observational research design where tweets or customer behaviour on twitter was closely observed without influencing them in any way. Data was collected via the twitter API and stored in memory for analysis. Python and the twitter API were used as research instruments and the target population were Econet , ZOL and Telone customers that use Twitter to engage with Econet. The sample size was made up of 200 tweets in one request and 3000 successive requests for older tweets. Data was analysed using python and associated libraries. The customer tweets were presented as a data frame. The study revealed that sentiment analysis provides a non-intrusive way of measuring customer satisfaction. The results for all telecommunications providers suggests that overall twitter engagement is positive with further study required to drill down on specific content and reveal hidden meanings that might have been missed by the sentiment analysis algorithm.

Acknowledgements

Special praise to God for the health and endurance to undertake this project and complete it to the satisfaction of Chinhoyi University of Technology. Also, to acknowledge the contribution of Mr H Mazhokota my supervisor who has vast industry experience within the telecommunications, banking and insurance sectors as a data science professional for his tremendous guidance and support, which added substance to the work. Also, I extend my appreciation to Master of Science in Data Analytics class of 2019-2020 especially the invaluable information I received from the discussions we had with those working within the Zimbabwean telecommunications customer service departments which has been valuable in shaping the facts in my dissertation.

Dedication

This work is dedicated to my daughter Myah Chamambo who watched her father working from home alternating between work and study during the December 2020 to April 2021 dissertation semester.

Table of Contents

List of Figures

Table of Tables

Abbreviations/Acronyms

POTRAZ – Postal Telecommunications Regulatory Authority of Zimbabwe
CES – Customer Effort Score
NPS – Net Promoter Score
CSAT - Customer Satisfaction

Chapter 1

1.0 Introduction

The most common approach used by most businesses to measure customer satisfaction is to conduct surveys. The trick to calculating it is to ask the right questions of your customers. The answers may not show you the correct data if you do not ask the right questions. You will not be able to recognise and fix areas of change in your company without the right data. Generally, the traditional way of gathering data poses a challenge to businesses since it tends to be biased. There are varying methods of collecting customer data and most methods rely on having an audience with the customer by sending them a survey or a form to complete. How does a business measure if the responses to the surveys where unbiased ? , they will not be able to tell the difference. This chapter describes the context of the research issue, the research goals, importance, scope, and limitations of the research on customer service assessment using sentiment analysis.

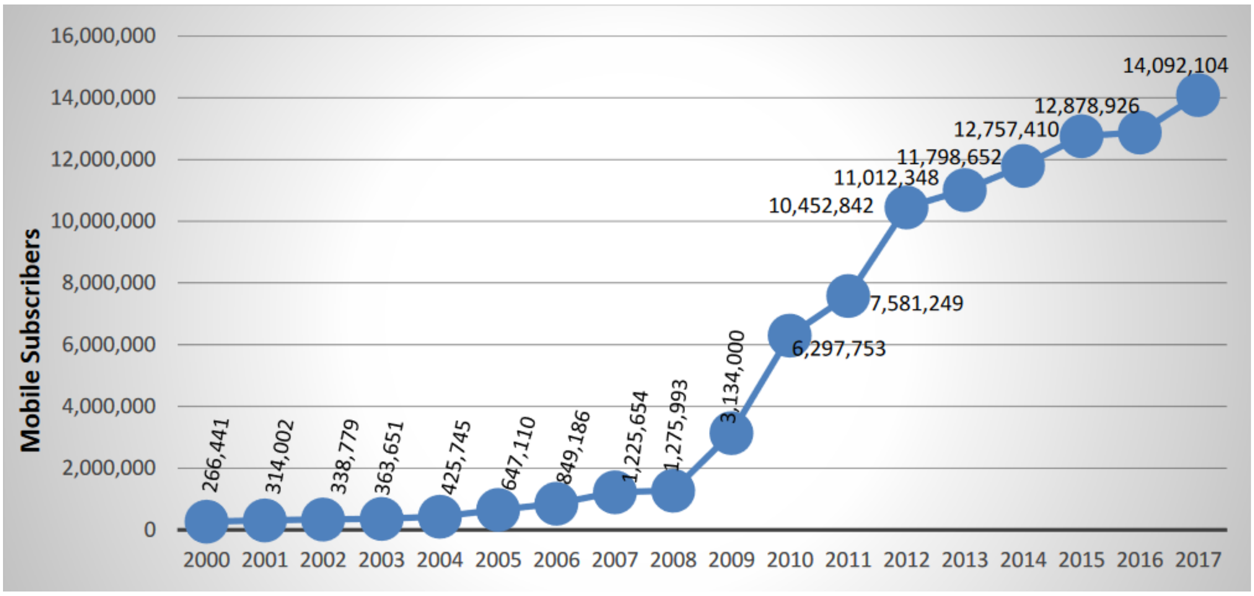
1.1 Background to the study

Customers' experiences have been described as a critical component to improving company efficiency (Makudza, 2020). In the services industry, where the quality of a service is measured by the strength of the service engagement, the need for customer relationship management has become more pronounced (Cajetan, 2018; Zeithaml et al., 1990). Given the increasingly large number of consumers using social media, most businesses globally have embraced social media to engage with current and prospective customers. These social networks have given them the ability to communicate with consumers in real time. There are using social media monitoring techniques to gauge how people are talking about their brand, products and services online. Because information spreads fast in social media, most organisations have strategically positioned their customer service departments to utilise social media since it helps them identify issues early, monitor and predict their growth. Customer service centres strive to improve customer experiences across the customer journey, from evaluating a product purchase to the after-sales support needed. The need to ensure customers are happy is an ongoing priority for any business that wants to survive. Mobile and fixed internet service providers in Zimbabwe have been focused more on expanding their network reach for the past 5 years but now they have shifted their focus to customers by improving and enhancing their products and services evidenced by various customer marketing initiatives that have been carried out by Econet and Netone through the buddie and One Fi marketing campaigns.

1.1.0 Zimbabwe mobile internet service providers – Subscribed Users

Although the sector's success is influenced by the economy, the mobile telecommunications market has grown significantly over the last two decades.
The sector is influenced by the economy in terms of service production and use, operational costs, expenditure, and so on. According to the annual sector results survey (Potraz, 2017), demand for mobile telephone services has increased. The graph below depicts the increase in active cell phone subscribers in Zimbabwe from 2000 to 2017:

Figure 1 - Surge in active mobile subscribers.



As of December 31, 2017, there were 14,092,104 registered mobile subscriptions. This is an increase from the 12,878,926 users that were counted in 2016.According to POTRAZ, an active mobile subscription is a mobile line that has sent or received a call, sent or received a letter, and/or used the internet at least once in the previous 90 days. The annual difference in active and total mobile subscribers as seen in Table 1:

Table 1 - Active Mobile Subscriptions (period 2016 – 2017)

Operator	Active Subs 2016	Active Subs 2017	% Change
Econet	6,360,904	7,488,588	17.7%
Telecel	1,805,612	1,646,411	-8.8%
NetOne	4,712,410	4,957,105	5.2%
Total	**12,878,926**	**14,092,104**	**9.4%**

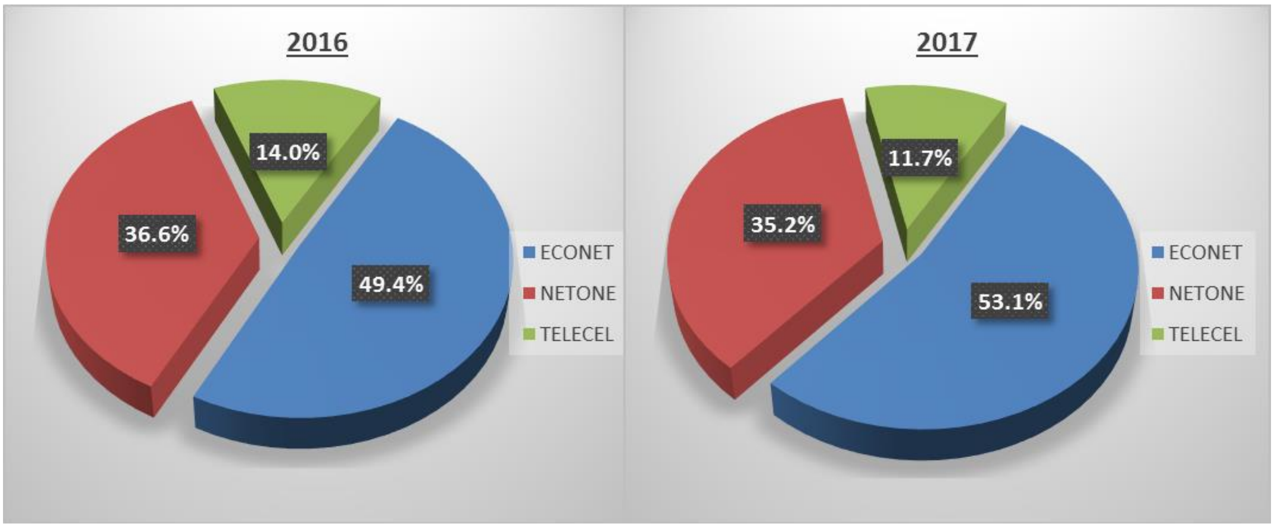
As shown above, Telecel was the only mobile operator to register a decline in active mobile subscriptions. Econet recorded the highest growth in active subscriptions of 17.7%.

Table 2 - Active Mobile Subscriptions (2020)

Operator	2nd Quarter 2020	3rd Quarter 2020	Variance (%)
Econet	8,766,968	8,603,084	-1.9%
NetOne	3,306,173	3,455,277	4.5%
Telecel	725,157	725,424	0.04%
Total	12,798,298	12,783,785	-0.1%

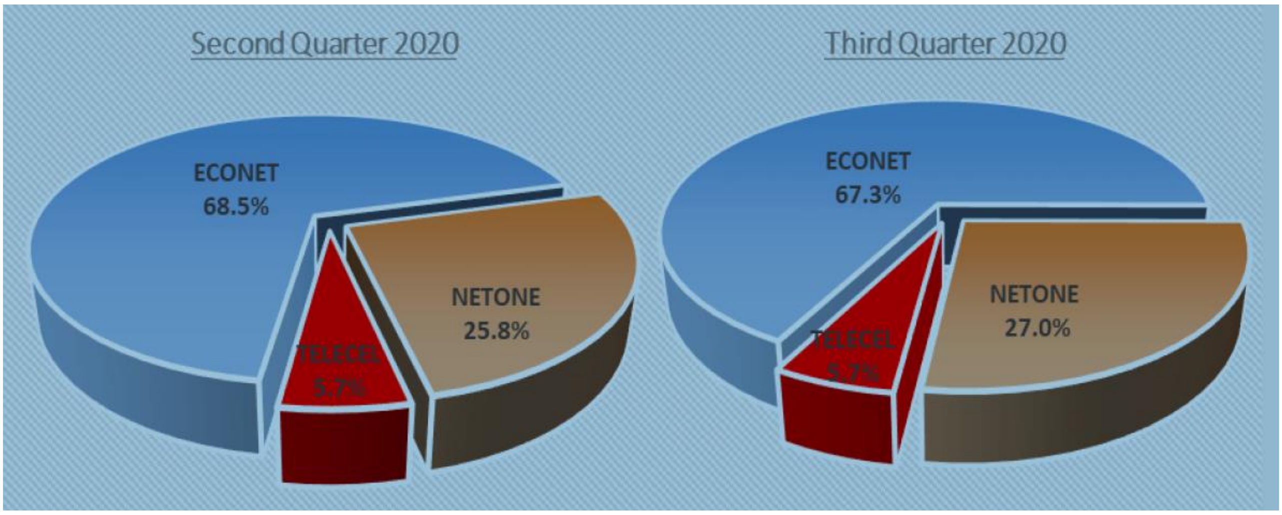
Active mobile telephone subscriptions declined by 0.1% to reach 12,783,785 as of 30 September 2020, from 12,798,298 recorded as of 30 June 2020. This however was because of the effects of COVID 19 lockdown restrictions in the formal and informal sector on disposable income.

Figure 2 - Market share of mobile subscriptions for 2016 and 2017



Econet continued to dominate the market with a market share of 53.1% at the end of 2017. Econet`s market share increased by 3.7% in line with the 17.7% growth in active subscriptions. On the other hand, Netone and Telecel lost market share by 1.4% and 2.3% respectively.

Figure 3-Market share of mobile subscriptions for period 2020

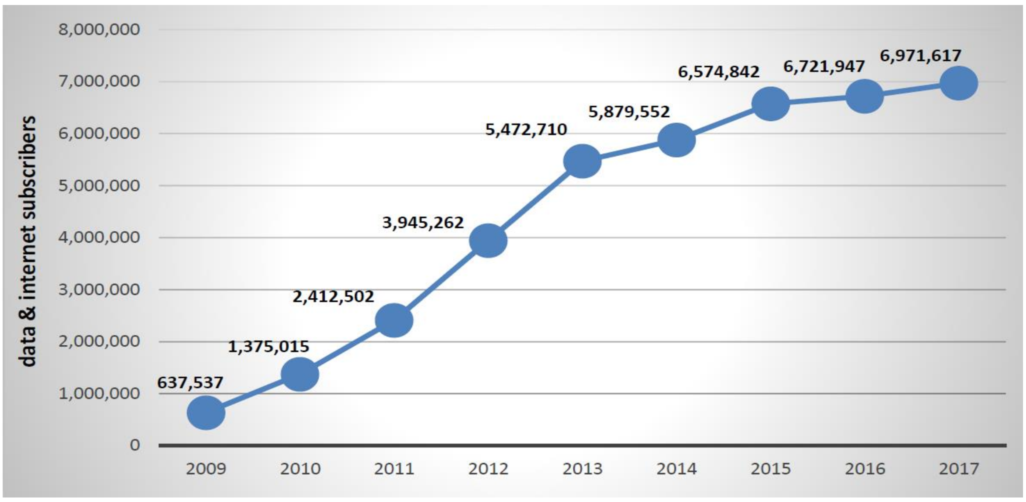


As shown above, Econet lost subscriber market share by 1.8% whereas NetOne gained market share by 1.8%. Telcel's market share remained unchanged at 5.7%.

1.1.1 Zimbabwe mobile internet service providers - Data Subscriptions

The total number of active internet subscriptions increased by 3.7% to reach 6,971,617 from 6,721,947 recorded in 2016.

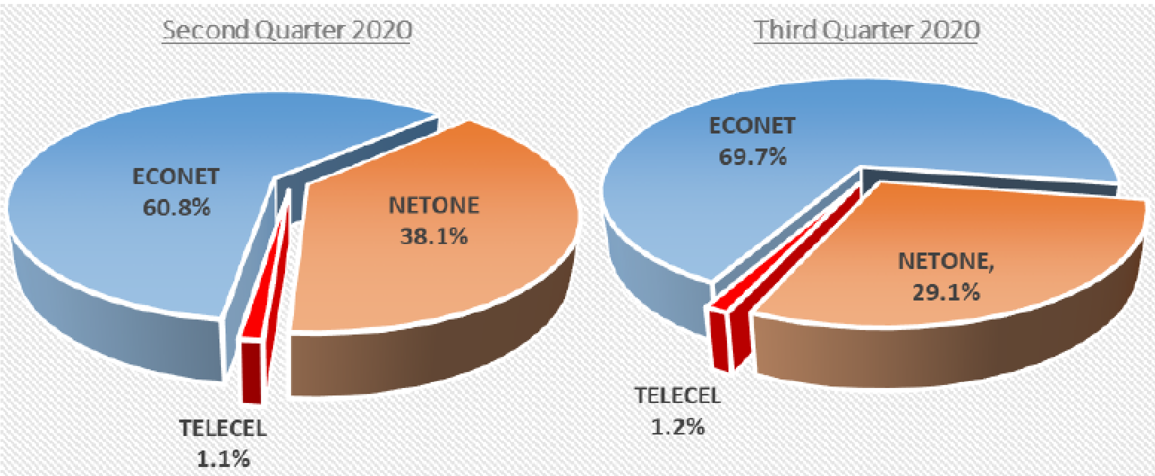
Figure 4 - Annual growth of internet subscriptions per category



Subsequently in 2020, Active Internet and data subscriptions grew by 5.6% to reach 8,726,904 from 8,267,268. As a result, the Internet penetration rate increased by 3.2% to reach 59.9% from 56.7% recorded in the previous quarter. Mobile Internet and data grew by a 43% to record 14,878 Terabytes from 10,407 Terabytes (TB) recorded in the second quarter of 2020. All the mobile operators recorded growth in Internet and data usage (POTRAZ ,2020).

The mobile telecommunications industry in Zimbabwe is dominated by three players namely, Econet Wireless, Telecel Zimbabwe and NetOne. The table below illustrates the market share of each of the three companies according to According to (POTRAZ ,2020) performance report.

Figure 5 – Mobile service provider market share



As shown above, Econet and Telecel gained 8.9% and 0.1% market share respectively, whereas NetOne lost market share by 9%.

Econet and Telecel acquired 8.9% and 0.1 percent market share, respectively, according to (Potraz,2020), while NetOne lost 9% market share. The three service providers have seen a lot of competition in the market contributing positively to the development of the telecommunications sector in Zimbabwe according to (Marumbwa and Mutsikiwa, 2013), with the operators aggressively competing for the market share in a bid to offer differentiated products. Most subscribers are subscribed to all three mobile providers, switching between one provider to another depending on use case. All 3 mobile service providers have been facing a decrease in their customer retention rates. Price was the differentiating factor which caused consumers to prefer and maximise on the cheaper promotional bundles instead of out-of-bundle data rates. However, customers are now looking for much more than just the price or the quality of the service. There is a growing demand across the industry to evaluate and improve customer experience across a product or service lifecycle. Businesses around the world are working on improving consumer service to minimize turnover, raise sales, and eventually become sustainable. A customer who has a good relationship with a company is most likely to return and become a lifelong customer. Mobile telecommunications providers in Zimbabwe have all adopted various customer engagement channels such as social media and emails to bridge the communications gap between customers and their customer services departments. Customer experience is essential to a company's competitiveness, according to business executives (McCall 2015), and marketing experts refer to it as the "fundamental foundation for marketing strategy" (Homburg et al. 2015; Lemon and Verhoef 2016). Due to the competition between service providers most notable Netone Pvt Ltd and Econet Wireless, POTRAZ realised that offering products and services alone wasn’t enough and through the POTRAZ Act [Chapter 12:05] which defines the Functions and Powers of the Authority , “To protect the needs of subscribers, buyers, and other users in terms of the standard and variety of postal and telecommunications facilities offered”, POTRAZ defined a framework which enabled them to collect metrics from telecommunications providers in order to determine the quality of service. According to the literature, no matter what service or product a customer buys or receives, the customer will have an experience; positive, negative, or indifferent, i.e., a service often arrives with an experience (Carbone and Haeckel 1994), and that all service experiences provide an opportunity for emotional interaction, no matter how tedious the product or service may be (Berry and Carbone 2007, Voss and Zomerdijk 2007)

1.1.2 Zimbabwe Data and Internet Service by technology

Active Internet and data subscriptions increased by 5.6 percent in the third quarter of 2020, to 8,726,904 from 8,267,268 in the second quarter. As a result, the Internet penetration rate rose by 3.2 percent to 59.9% in the third quarter, up from 56.7 percent the previous quarter.

Table 3 - Active data and Internet subscriptions by technology.

Active Internet Subscriptions Technology	2nd Quarter 2020	3rd Quarter 2020	Variance (%)

3G/HSDPA/LTE	8,081,986	8,547,844	5.8%
Leased Lines	2,056	1,646	-19.9%
DSL	110,981	110,586	-0.4%
WiMAX	1,927	1,953	1.3%
CDMA	21,080	15,313	-27.4%
VSAT	2,661	2,682	0.8%
Active Fibre subscriptions	46,577	46,880	0.7%
Total	**8,267,268**	**8,726,904**	**5.6%**

As seen above, the three Internet and data groups with negative growth were leased cables, DSL, and CDMA. Mobile Internet subscriptions grew at a faster pace than fixed Internet and data subscriptions in the period under study.

1.1.3 Zimbabwe fixed internet providers - International Internet Connectivity

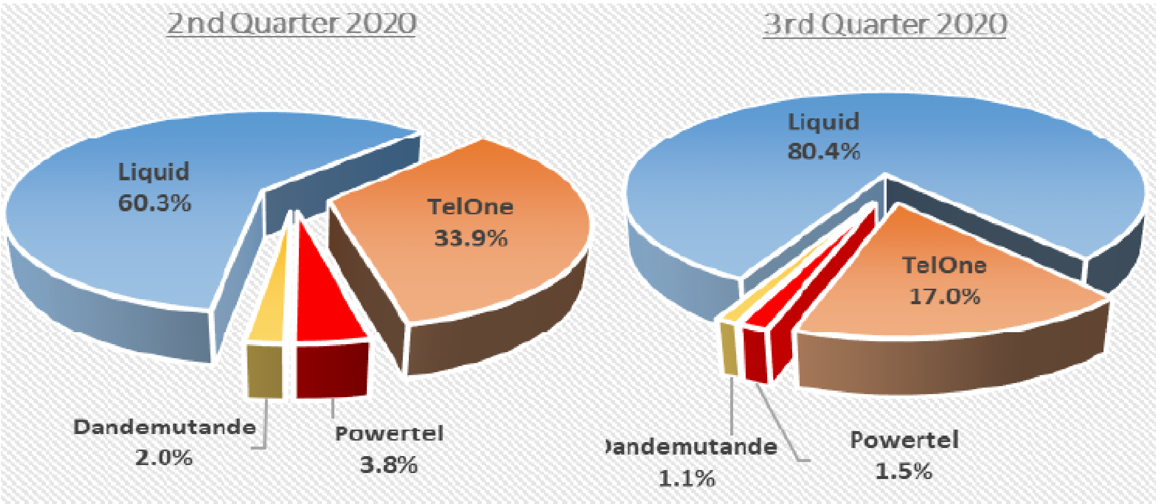
Following expansion by Liquid, Telone, and Dandemutande, the third quarter of 2020 saw a massive increase in fitted international Internet bandwidth, as seen in the table below:

Table 4 - International Internet capacity period of 2020

Equipped International Internet Bandwidth Capacity	2nd Quarter 2020	3rd Quarter 2020	Variance (%)
Liquid	80,000	260,000	225.0%
TelOne	45,000	55,000	22.2%
Powertel	5,000	5,000	0.0%
Dandemutande	2,715	3,416	25.8%
Total	**132,715**	**323,416**	**143.7%**

IAPs will need to begin expanding their bandwidth capacities to satisfy demand as demand for Internet and data grows due to the adoption of e-learning, telecommuting, and e-health, among other things. As seen in Figure 6 below, Liquid gained market share in comparison to other Internet Access Providers because of a massive increase in equipped international Internet bandwidth capacity:

Figure 6 - Market Share of International Internet Bandwidth Capacity



In the third quarter of 2020, used international incoming bandwidth capacity rose by 16.8% to 188,093Mbps, up from 128,173Mbps in the second quarter. In the third quarter of 2020, used international outgoing bandwidth capacity rose by 4.2 percent, to 40,033Mbps, up from 38,428Mbps in the second quarter. The capacity of used international incoming bandwidth per IAP is shown in the following table

Table 5 - Used Incoming International Internet Bandwidth Capacity

		Used international outgoing bandwidth capacity
Liquid	89,072	26,528
TelOne	55,000	9,475
Powertel	3,410	3,410
Dandemutande	2,183	620
Total	149,665	40,033

Because of the rising need for data and Internet services, as well as the growing availability of local web content, used international Internet bandwidth capacity is projected to continue to rise.

This oversight from POTRAZ has provided telecommunication companies an opportunity to improve their customer experience processes. Consumer experience (CX) is characterized as a customer's reaction to experiences with an organization before, during, or after transaction or consumption, through multiple channels and time, according to (Holmlund et al., 2020). Customer satisfaction has emerged as a long-term source of competitive differentiation among mobile carriers. Telecommunication companies have been trying to quantify the impact of customer experience on their revenues. A customer's relationships with their service provider include bill payment, fixing a technical problem, and updating their service. The telecommunications industry has realised significant growth since its inception and during its infancy when telecommunications providers were expanding their coverage, the quality of service was so poor with customers experiencing call drops, poor call reception and slow internet speeds. While service quality is one of the metrics used to retain customers, it was not long since the oversight from POTRAZ helped improve service quality across the telecommunications service providers. This forced the telecommunications providers to introduce various techniques that would help them retain and gain new customers in the process. We saw Econet introducing a myriad of marketing campaigns in the form of their flagship product “buddie” while on the other hand Netone also introduced “easycall”. Telecommunications providers have been switching between one form of campaign to another , hiring more staff for their customer service departments in a quest to improve customer experience. These methods have limitations. Text mining and other new innovations have the ability to improve consumer service measurement and management (Keiningham et al. 2017; Lemon and Verhoef 2016; Verhoef, Kooge, and Walk 2016; Villarroel Ordenes et al. 2014). Researchers have been searching for ways to obtain valuable insights from big data produced during customer service journeys, as well as to understand, process, and analyze it , and improve customer experiences. Many organisations focus primarily on their own perspective on what customer experience is. We depart from this by focusing on the customer and demonstrate that customer experience analytics that use big data strategies to the customer experience will provide valuable insights. We draw on seminal studies in meaning formation by McColl-Kennedy et al. (2012) with later work by Macdonald, Kleinaltenkamp, and Wilson (2016) that emphasizes the significance of relationships at touchpoints and meaning, foundational work by Verhoef et al. (2009), and later work by Lemon and Verhoef (2016) that emphasizes the importance of both emotion and cognition. The value of understanding the consumer experience as a journey of many touchpoints over time is emphasized.

The internet has evolved into a vast repository of skewed information.
Twitter and other social media platforms are thought of as online diaries where millions of people share their feelings and thoughts through their everyday interactions (Carvalho, Prado, Plastino, 2014). Sentiment analysis is one of the most critical foundations of social media analytics, as well as one of the most successful social media tracking strategies used by creative businesses to remain competitive. Because of their market presence in Zimbabwe, mobile telecommunication service providers such as Econet and Netone stand to gain further from the amount of user-generated content on social media. Telecommunication companies in zimbabwe play a very important role of providing communication services to most businesses, connecting them to the rest of the world through the internet. Given the number of combined subscribers these providers have, making use of social media analytics in determining customer satisfaction will provide them with deeper insights on how they can tailor their packages according to their customers and in the process position their brand competitively. Customer engagement exercises globally have been influenced by developments in social media analytics. Customer interaction refers to the degree to which a person participates in and is connected to an organization's services and/or events, which are initiated by either the customer or the organization (Vivek, Beatty, & Morgan, 2012). In this digital era of technology where information is easily accessible, social media sites such as Facebook and twitter has generated vast amounts of data. According to (Kaplan and Haenlein ,2010) Social networking is described as a collection of web- and mobile-based Internet applications that enable the development, access, and sharing of ubiquitously available user-generated content. Businesses globally now rely on informed customer experience processes to grow their business portfolios and get a comparative advantage on their competitors.

In the period 2016 to-date, the Zimbabwean mobile telecommunication service provider space continued to see POTRAZ, the telecommunications regulator in Zimbabwe playing a pivotal role , policing and standardising of key performance indicators across all providers. This has resulted in customers switching between one provider to the other in a quest for the best provider from service quality to customer service. In the process mobile providers have introduced new ways of marketing their products and servicing their customers. Mobile providers are now scrambling for the same clients who still see value in the products and services.

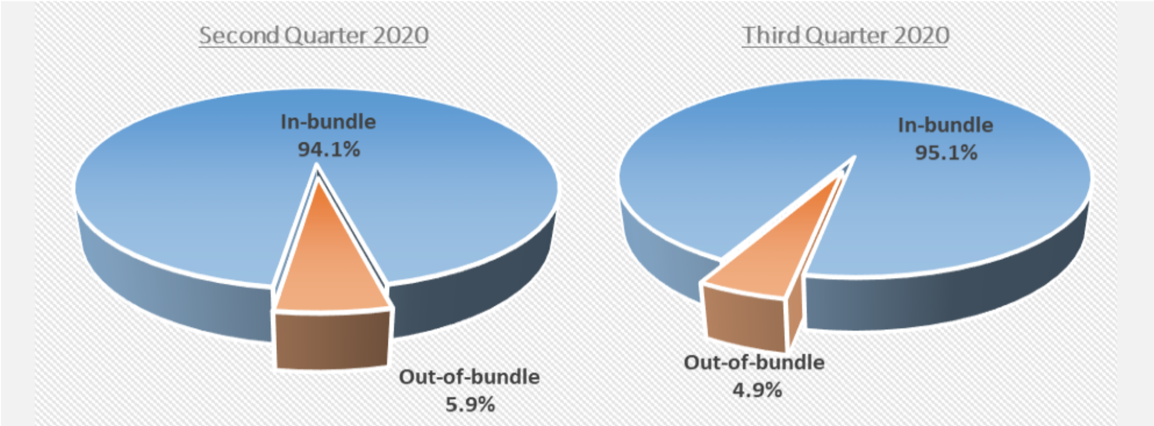
Zimbabwean businesses do not have the means and skills to process large amounts of data let alone unstructured data being generated on social media. The uneven economic level playing field in zimbabwe calls for businesses to think outside the box when it comes to embracing new market research methodologies. Some innovative companies have embraced social media market research which has brought in a new perspective of product and service awareness to customers.

In this study, the researcher is going to look at how Zimbabwean businesses can incorporate twitter sentiment analysis in their market research process. According to (Komariah, Machbub, Prihatmanto and Sin , 2016) , Twitter, as a social media data provider, has a significant impact on fast and responsive responses to political, social, and economic problems, as well as playing a key role in shaping consumer perceptions of a company's brand or product. It can be used to predict brand or product perception based on people’s sentiment.Traditional market research methods have been affected by changes in consumer communication patterns, (Patino, A., Pitta, D.A. and Quinones, R. 2012).

1.2 Statement of the problem

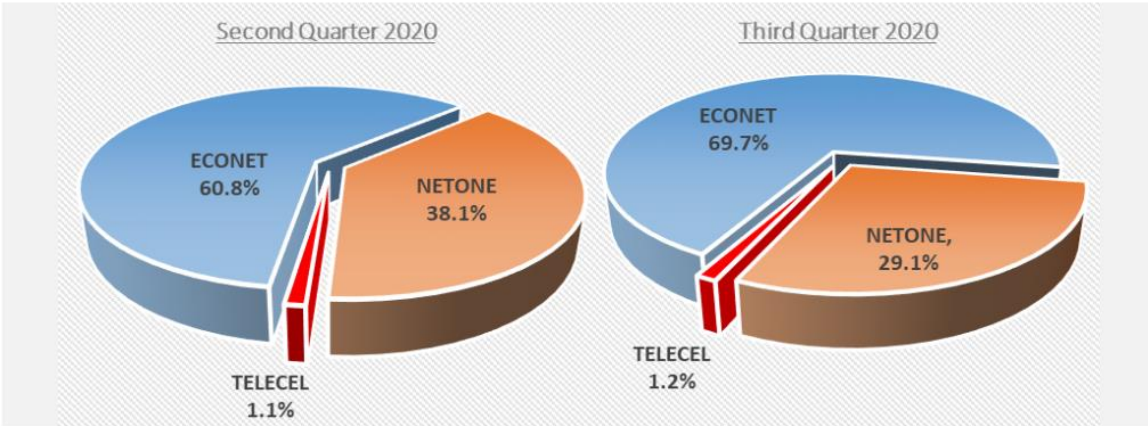
There has been a resurgence in mobile internet subscriptions since the COVID-19 outbreak in Zimbabwe. Businesses operations around the world got affected with Governments in various countries responding to the effects of the pandemic by enforcing lockdowns. Organisations in Zimbabwe began to adopt teleworking to maintain business operations. POTRAZ believes the growth will continue because of the adoption of e-learning, telecommuting, and e-conferencing. This new operating model has caused a strain in customer service departments within mobile service providers. The COVID-19 pandemic and subsequent lockdown measures put in place in Zimbabwe has highlighted the significance of telecommunications with official statistics from the country’s industry regulator reflecting an increase in mobile internet and data.

Figure 7 - Mobile Data & Internet usage



According to a survey by POTRAZ (Potraz, 2020), mobile Internet and data increased by 43% to 14,878 Terabytes (TB) from 10,407 TB in the second quarter of 2020.
The use of the Internet and data has increased across the board for all telecom operators. In-bundle data accounted for 95.1 percent of gross mobile Internet and data use in the second quarter of 2020, up from 94.1 percent in the previous quarter. With people stuck at home for months, mobile telecommunication providers have gained from those working remotely and those posting on social media. The telecoms sector has an advantage in that under the various lockdown levels, it continued to operate. The COVID-19 pandemic has illustrated the vital role of telecommunications networks in maintaining companies, states, and communities linked and operational. By delivering business-critical connectivity and flexibility, promoting work-from-home plans, and keeping individuals and communities engaged and alert with access to basic resources during mandated social isolation, the sector has been critical in keeping the global economy going during lockdown. As a result, as seen in Figure 8, many telecom players who provide broadband have benefited from an increase in data traffic. If more companies embrace teleworking, data and Internet services would accelerate market growth. Figure 8 shows a quarterly analysis of the market share of mobile Internet and data usage:

Figure 8 - Market Share of Internet& Data Traffic



Econet and Telecel gained 8.9% and 0.1% market share respectively ,whereas NetOne lost market share by 9%.

Mobile service providers in Zimbabwe are product-centric rather than customer-centric which makes departments operate in silos. In a product centric approach, marketing only focuses on marketing while operations focus on operations . The surge in mobile internet traffic has affected current customer experience processes. Given the fluctuation of active subscribers, subscriber retention would be critical for operators in the future. The different service industries are projected to become more competitive, with mobile service carriers competing on devices, service offers, and pricing. POTRAZ's role as the regulator introduced measures which forced customer executives to shift their focus towards customers and more deliberately on customer experiences. Customer service departments have been forced to restructure their teams which is affecting the timely resolution of tickets raised by customers.

1.3 Existing customer experience evaluation approaches

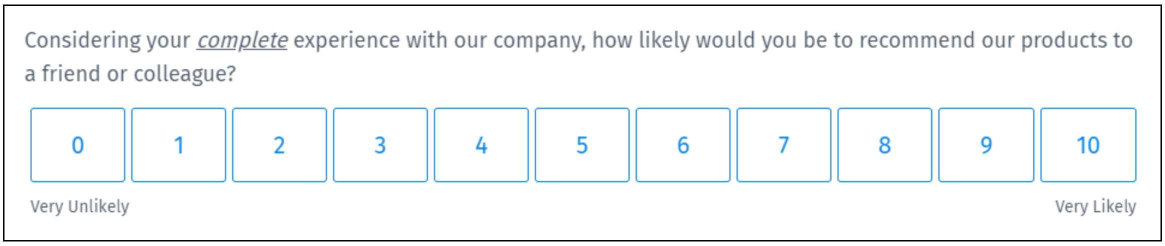
Currently, the only way for mobile service providers to learn how customers respond to a product or service is to hold focus groups or send out surveys, all of which offer no chance of yielding a statistically meaningful number of responses. In today's dynamic environment, businesses who fail to meet their consumers' expectations and offer value to them risk losing them to other service suppliers that can meet their needs (Gate, Tatenda). The quantitative feedback you receive from customers in terms of NPS, CSAT or CES metrics is valuable in terms of understanding how your business is performing on average but does not give you a true picture. Examining customer feedback in terms of verbatim comments, whether in a survey, on social media can be a time consuming and oftentimes the negative feedback takes priority. These traditional methods of evaluating customer experience and measuring customer satisfaction are not as effective. Sentiment analysis can be used to make a massive difference in how you interpret and understand customer feedback.

Most businesses in Zimbabwe are not technology enabled and the new operating model has forced them to rethink their information technology strategies through digital transformation initiatives. The telecommunications industry has not been spared. Teleworking has emerged as the ideal operating model for Zimbabwean mobile telecommunications providers in the year 2020 to 2021 due to government enforced lockdowns. No one was ready and in turn , this has affected their customer base significantly due to the inadequacy of existing customer experience strategies. This shortcoming has severely affected their business operations since employees are now required to work from home. The number of registered customer service tickets has risen dramatically. Mobile telecommunications providers such as Econet and Netone have enabled businesses to enable their employees to continue to work remotely by providing internet connectivity. The surge of customer support tickets being experienced by mobile providers has exposed a huge gap in existing customer services processes. There are finding it difficult to evaluate customer experience and measure customer satisfaction due to the number of connected individuals and the number of raised support tickets. Customers and businesses are now engaging each other on various social media platforms. Focusing on twitter, we see all 3 mobile providers Econet, Netone and Telecel engaging their customers online through their twitter handles. Customer support agents are now re-aligning current procedures with the new standard, figuring out how to cope with increased ticket demand and respond quickly to urgent questions. The effects of lockdown on mobile service provider cashflows has left them struggling to grow their portfolios while maintaining their existing customer base. Since the initial lockdown in Zimbabwe, which was enforced in March 2020, there are now focusing on maintaining existing customers while at the same time developing modern customer engagement processes. Any company now realizes that if they wish to keep their clients, they must invest in their experience. Most businesses have been focusing on rebuilding their customer services departments by hiring more employees to assist customers. This is meant to retain existing customers while acquiring new ones. On top of hiring new employees, organisations have been using standard customer services metrics to evaluate customer experience and measure customer satisfaction. There are 6 popular customer experience metrics and KPIs that are currently used by organisations to track customer experience, Net promoter score (NPS), Customer satisfaction (CSAT), Churn rate, Retention rate, Customer lifetime value (CLV), Customer effort score (CES). Customers are responding to surveys in decreasing numbers, resulting in too small sampling sizes and statistically meaningless results. All these metrics have one thing in common, customer satisfaction is measured when seen from the eyes of the service provider .The customers you already have to some extent but will not give you deeper insights of what the customers are saying about your products and services. These traditional methods are slowly getting replaced by developments in Big data analytics. The instability of the Zimbabwean economy makes it even worse; a customer will switch between one service provider to another depending on what there are looking for. A change in the operating environment means telecommunication companies should respond accordingly but there are not equipped enough to understand this customer behaviour. To counter the changes in the operating environment, some telecommunications companies have Incorporated social media analytics such as Google analytics or Facebook ads in their customer engagement initiatives. While this is a worthwhile effort, twitter sentiment analysis goes a step further and provides businesses with more granular detail into what your audience are really saying about their product or service. Consumers will always want the best products and services. However, for business owners, these are not the only things you should focus on. Customer satisfaction is something companies of all industries should work on. It is considered as a vital component of any business. After all, it could make or break your company. Learning about how your customers perceive your business and how happy they are with your products and services is one step to achieving and/or maintaining customer satisfaction. Below the researcher takes a closer look at how existing methods of measuring customer satisfaction work

1.3.1 Net promoter score (NPS)

NPS is a management tool that can determine the customers' loyalty to the company. It measures customer perception based on a key question, "How probable are you to recommend this firm's product or service to a peer on a scale of 0 to 10?"

Figure 9 - Net promoter score scale



The categories of customers are categorized using the scale.

Figure 10 - Net promoter score scale

1. **Detractors**: Customers who give a company a score of 0 to 6 fall into this group. They are not trustworthy and are more likely to spread derogatory information about your business.

2. **Promoters**: Customers who give a company a 9 or 10 on a scale of one to ten are the most loyal customers. They enthusiastically spread the word about your company to their friends and families.

3. **Passives**: Customers who give a company a score of 7-8 fall into this category. Passives are in the middle of the two categories mentioned above. They will not enthusiastically promote your brand to anyone, but they will not prevent their friends and family members from doing so.

1.3.2 NPS formular used by most organisations.

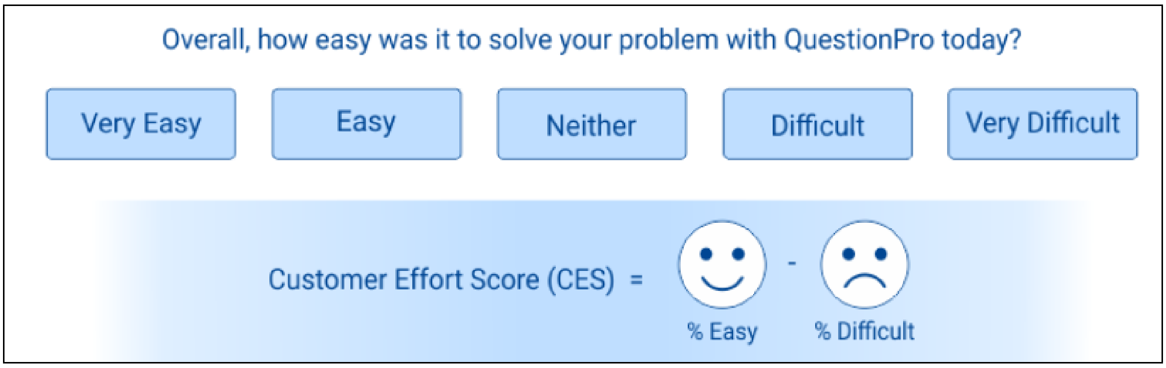
[[**Number of promoters** **−** **Number of detractors**] **/** **Number of responses**] **X** **100**

Subtract the percentage of detractors from the percentage of promoters. The NPS ranges from -100 to 100, based on how consumers perceive a company's brand and how committed they are to it. Although NPS provides information on where the company stands, it does not explain why the score is what it is. After the NPS query, most businesses add follow-up questions to elicit additional information from their clients.

1.3.3 Customer effort score (CES)

CES is a metric that helps organizations know the client's ease of their customer experience. CES surveys ask customers to rate the ease of using a brand's product or services on a scale of "very difficult" to "very easy." Some products or services can be difficult to use and demand a considerable amount of work on the part of the consumer. The aim of CES surveys is to reduce commitment while increasing loyalty.
Depending about what the organization hopes to accomplish through the report, there might be follow-up questions. When the commitment needed to use a service or product increases, loyalty declines. The aim is to make the process as simple as possible.
If customers are increasingly reporting events, it impacts their CES .This increases chances of subscribers switching over to other service providers. Being attentive to clients, proactively assisting them, and anticipating their desires and requirements would be beneficial.

Figure 11 - Customer effort score



1.3.4 Customer lifetime value (CLV)

Customer Lifetime Value is a measure of a customer's worth to an organization over time. A simple Customer Lifetime Value model that is used by most organisations is shown by the formula below.

[**Annual profit contribution per customer**] **X** [**Average number of years that they remain a customer**] **–** [**the initial cost of customer acquisition**]

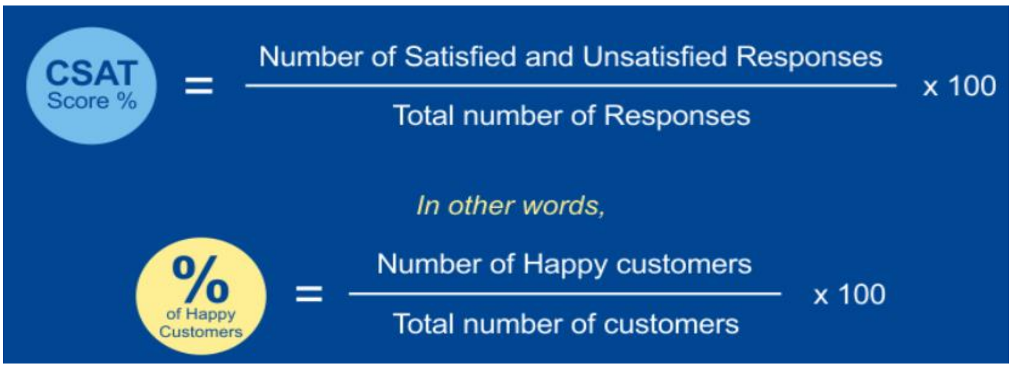
1.3.5 Customer satisfaction (CSAT)

CSAT is a basic measurement of a customer's satisfaction with a brand's product, services, and other offerings. It is usually used by brands to gauge the customer's satisfaction at key interaction times. CSAT respondents can rate their satisfaction on a scale of 1 to 5 as follows:

very unsatisfied (1), unsatisfied (2), neutral (3), satisfied (4), very satisfied (5).

The Customer Satisfaction Score tells you whether or not your clients are satisfied. The Customer Satisfaction Index (CSAT) is determined by dividing the number of satisfied customers by the total number of customers. Customer satisfaction surveys, which are typically performed at the point of contact, provide very credible and precise information. The responses are sincere and candid, and the response rates are satisfactory.

Figure 12 - Customer satisfaction (CSAT)



1.3.6 Churn rate

A certain number of a company's customers will still leave without leaving a review or getting input. The customer churn rate (CCR) is the number of consumers who leave a company over time. When a company's overall unique survey answers are compared to the number of consumers who left, the result gives an estimate of how many people left without leaving suggestions. In the vast majority of instances, it is fair to assume that these consumers were often dissatisfied with their experience. It is important to keep loyal consumers because acquiring new ones will cost up to seven times more. Monitoring CCR allows you to spot any changes that could have an effect on your business and take action to reduce churn.

1.3.6.1 Calculating churn rate

To calculate CCR, first choose a measurement time, such as a year, and then deduct the number of customers at the end of the year from the number at the beginning. CCR is calculated by dividing this figure by the number of consumers at the beginning of the year.

Figure 13 - Churn rate

$$\text{Customer Churn Rate} = \frac{\text{No. Of Customers lost}}{\text{Total no. of Customers (Period)}} \times 100$$

CCR measurement is not the last step. Once the data has been collected, organizations must search for additional data that is responsible for it and determine what preventative steps can be done to reduce it.

1.3.7 Customer reviews

Although this covers all the bases, you should be aware of any comments or reviews you get from portals, blogs, or social media platforms. It is impossible to overestimate the value of ratings. According to a new survey, product ratings are valued by over 90% of shoppers over product details.

Figure 14 - Customer reviews

Good ratings and tips on websites or portals that your future customers can visit are always welcome. You may ask them to write a review if they have favourable comments or suggestions about your goods or services. If they are pleased and fulfilled, they would gladly recommend your business to their friends and colleagues.

1.3.8 Social Media Monitoring

The effect of social media on business-customer relationships has been enormous. Previously, a good or bad service experience would only be shared with immediate families and friends, but social media provided a platform and future reach to millions of users.As a result, it's the ideal way to find out what your clients actually think of you. That is, if you have the right resources to keep track of it. Of course, Facebook and Twitter are essential to track, but so are websites like Quora, Yelp, TripAdvisor, and others.Twitter offers telecommunications providers extensive amounts of real time public data which can be used to track customer experience. The sentiment predictive model will enable them to engage their audience gathering insights about what the audience is saying about their product or service. This empowers businesses to participate in social media listening instead of just monitoring. Social media listening then allows businesses to influence the outcome of any customer engagement by choosing specific words that will bring forth a more positive sentiment.

1.4 Research objectives

Studies on sentiment analysis have been carried out by many researchers in the field of market research and customer services. This research focuses on how mobile and fixed internet service providers can use sentiment analysis to evaluate and measure customer satisfaction. Sentiment analysis helps organizations with product, service and brand validation. It gives organizations quantifiable data about customer perceptions. This informs campaign decisions through the evaluation of customer sentiment.

The study aims to achieve the following objectives:

1. To measure customer satisfaction using customer sentiment from tweets
2. To extract the most prevalent words that contribute to a negative or positive sentiment.
3. To identify words that can be used to improve customer engagement.
4. To develop a twitter sentiment predictive model

1.5 Research questions

The research sought to answer the following questions:

1. How do we measure customer satisfaction using sentiment analysis?
2. What words can be used to improve customer engagement?
3. How do we measure customer sentiment from words?
4. What is the general customer sentiment towards services on offer?

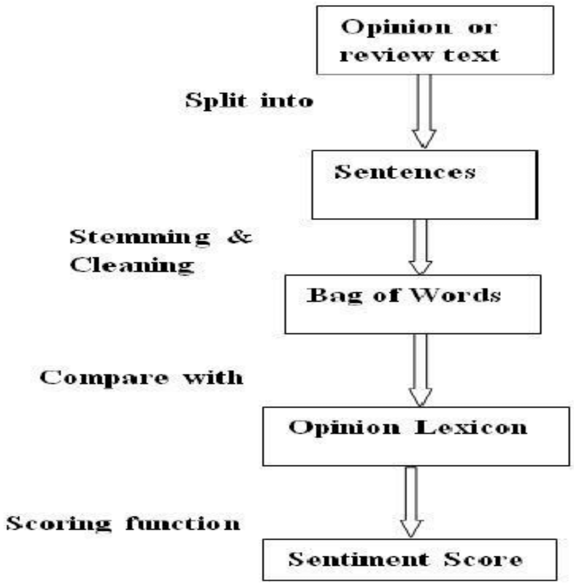
1.6 Research propositions

H1: Tracking customer satisfaction using sentiment analysis yields better results than using NPS or CSAT scores
H2: Analysed customer sentiments show a true picture of the quality of your service or product
H3: The effect of certain words during customer engagement influences how customers perceive your product /service

1.7 Conceptual framework

The researcher will adopt the following conceptual framework for this study. Textual data when extracted in its raw form based on previous studies should undergo transformations for it to be usable. The basic building blocks on sentiment analysis is being able to understand the meaning of words , sentences and the context that there are used. The researcher will quantify and analyse the presence, meanings and relationships of such words, themes, or concepts. The researcher can then make inferences about the messages within the tweets, the service providers and the customers there are engaging.

Figure 15 - Conceptual framework



Phase 1 - Data will be collected from Twitter using the Twitter Application Program Interface (API)

Phase 2 - Dataset will be pre-process (cleaning and removing noise)

Phase 3 - Classify text into positive, negative and neutral polarity.

Phase 4 – Sentiment calculation

1.8 Significance of the study

The study will help the researcher broaden his knowledge on how sentiment analysis can be used to evaluate and measure customer experience. The market being analysed will benefit from the analysis as well, as it will have an in-depth look into how they measure consumer loyalty. The study would also make recommendations on how to improve the model's implementation to achieve the expected results. The study would also provide the industry with more information on sentiment analysis and how it can be applied successfully. The study will also be used as literature by the academic community. The study will be accessible to anyone in the academic community who would like to know more about the subject. Customer satisfaction is fundamental to business growth.

The telecommunication industry in zimbabwe plays an important role in enabling business in other industries. Evaluating customer satisfaction through sentiment analysis provides cheaper and quicker results as compared to other alternatives NPS and CSAT. Findings from this study will benefit all telecommunication companies in incorporating twitter sentiment analysis in their customer experience programs. Sentiment analysis allows for a quick and accurate assessment of customer opinion in real time.

1.9 Delimitations of the study

There are multiple sources of customer data that can be augmented with social media data such as call records, data from incident management systems and market research data from surveys but the research was focused on data from twitter. This may not be sufficient for the purposes of generalisation (Yin, 2009). Many Zimbabwean businesses have embraced digital transformation initiatives mainly focusing on a widely spread online digital footprint. However, the study will be limited to mobile and fixed internet companies that have an online presence on twitter. Also due to time constraints, the model will not be multilingual. The model will not cater for Native Zimbabwean languages Shona and Ndebele within its language model. Data collected from tweeter that the researcher will encounter in the stream will not be stored to disk so that the researcher can refer to it and perform offline analysis. The decision to store the data in-memory during extracting and analysis forces the researcher to concentrate on the research objectives rather than the nuances of setting up storage servers which adds to the complexity of the architecture. Natural language processing (NLP) still struggles with the inability to distinguish between various senses of words and phrases, as well as recognizing sarcasm and sarcastic expressions. Although word uncertainty applies to the same word with different meanings, current approaches are still focused on the degree of similarity rather than the intended interpretation, resulting in poor functional implementations.

1. 10 Chapter summary

This chapter outlined the most common methods used to measure customer satisfaction with respect to how there are being used within mobile and fixed internet service providers. Various conventional data sources were explored with an introduction to social media analytics and how it can be used as a data source by businesses. The objectives of the study were stated emphasising how social media analytics ,specifically the application of sentiment analysis techniques can greatly improve customer experience and in the process provide the ability to measure customer satisfaction.

Chapter 2 Literature Review

2.0 Introduction

In this section, the researcher presents the review of previous studies done on various methods used to evaluate customer satisfaction. The work done by several researchers was reviewed in line with the objectives and the methodology of this study. The review was guided by already existing theories in customer satisfaction and sentiment analysis techniques. Literature on how organisations approach customer engagement exercises was reviewed. The popularity of social media was analysed particularly how it has necessitated organisations to find ways to tap into the vast amounts of data to improve customer satisfaction.

2.1 Related work – Sentiment analysis

Sentiment analysis is becoming more common. Current studies have relied on information extracted from social media. SVM, Naive Bayes, Decision trees, and other classifiers have also been used to predict the polarity of sentences. These classifiers were combined with word embeddings to yield cutting-edge performance.

Giachanou et al. published a detailed study on the most used methods for twitter sentiment analysis, such as SVM and Naive Bayes, as well as data sets like the Edinburg Twitter corpus, Stanford Twitter Sentiment, Sander’s corpus, and sentiment lexicons like SentiWordNet and MPQA. Da Silva et al. proposed hybrid models that combined SVM, Naive Bayes, Decision Trees, and other techniques to achieve 81.06 percent accuracy. A comparison of feature hashing and bag-of-words was also discussed, with the conclusion that feature hashing appears to be a reasonable choice when computational performance is important, but bag-of-words appears to be a better choice when accuracy is important and seems to be the most appropriate choice.

Hassan et al. suggested bootstrapping frameworks to ensure accuracy of performance with unbalanced data from Telco, Tch, and Pharma tweets. To generate models with greater precision in predicting sentences of extreme polarity, a two-stage method of expansion and contraction was used. Araque et al method .'s blends conventional surface models based on manual attribute extraction with deep learning strategies for sentiment analysis using word embeddings. For the F-measure test, six public data sets were used. According to a comparative analysis, the ensemble outperforms the baseline model, which uses only deep learning-based classifiers.

Muhammad et al. used the term frequency-inverse document frequency matching model to investigate the polarity of a sentence as well as the sentiment of the whole document with respect to the topic in a not-yet-explored Bangla text. Around 1500 Bangla comments from different social media platforms were used as the data collection. Precision and recall were used to assess the classifier's precision in predicting the polarity of a sentence, while accuracy was used to assess the classifier's overall results.

To characterize Tamil movie feedback as positive or negative, Se et al. used supervised machine learning techniques such as SVM, Maxent classifier, Decision Trees, and Naive Bayes. SVM outperformed all other models in terms of precision (which was estimated to be 75.9%) and cross-validation, with Decision Trees coming in second. Negative polarity was labeled 1 and positive polarity was labeled +1. The experiment was conducted using Tamil SentiWordNet and without it. The highest level of accuracy achieved with Tamil SentiWordNet was 4% higher than without it.

Uma et al. suggested a method for extracting information from Tamil and English tweets and categorizing each sentence as positive, negative, or neutral using the SVM classifier. The semantic similarity between words was determined by calculating path-length-based similarity, and WordNet was used to capture the context of a word.
The data collection was Twitter data obtained via the TwitterAPI. Weka was used to conduct tweet classification using SVM and a confusion matrix was obtained for evaluation. The pre-processing stage included using StringToWordVector. The F-measure obtained with SVM was 0.741, while the F-measure obtained with Naive Bayes was 0.68, indicating that SVM is a better option than Naive Bayes for sentiment classification of textual data.

Sentiment classification was conducted on the Sentiment Analysis of Indian Languages (SAIL) 2015 data by Phani et al.There were two groups of grouping experiments: (1) two-class: positive and negative classes, and (2) three-class: positive, negative, and neutral classes. On the training results, stratified 10-fold cross-validation was used. As compared to models from SAIL 2015, the models performed exceptionally well, with accuracy of 56.96 percent for Hindi, 51.25 percent for Bengali, and 45.24 percent for Tamil.

2.2 Customer satisfaction?

Customer satisfaction, according to Kotler and Armstrong, quoted in Karim and Chowdhury (2014), is a person's feelings of enjoyment or dissatisfaction as a result of a product's perceived success in addition to expectations. Similarly, Angelova and Zekiri (2011) describe consumer satisfaction as the outcome felt by those who have had their goals met by a company's results.The two concepts emphasize the importance of suppliers of products or services meeting customers' needs as a component of consumer satisfaction. As Naik, Gantasala, and Prabakar (2010) point out, one of the key goals of any company is to please clients, so retaining existing customers is more lucrative than acquiring new ones to replace those lost. According to McColl-Kennedy and Schneider (2000), cited in Naik, Gantasala, and Prabakar (2010), management and marketing theorists emphasize the relevance of customer loyalty for a business's success because of this realization. In studying customer satisfaction in the public sector and its impact on performance, Zamil and Shammot (2011) submit that customer satisfaction is critical for public sector organizations, contending that the customer needs services that satisfy him/her and equilibrate with his/her expectations. They further argue that if customer satisfaction is not achieved by public sector organizations, the customer will feel that his/her satisfaction is ignored, and this causes more complaints (Zamil & Shammot, 2011 Similarly, Chakraartyet al. (1996), as quoted in Karim & Chowdhury (2014), argue that continuous systemic assessment of customer satisfaction is needed because a happy customer is a real asset for an enterprise that ensures long-term sustainability even in a time of great competition. This viewpoint aligns with the International Social Security Association's assertion, in relation to social security organizations, that as individuals are given the ability to use the social security system in the manner that best serves their interests, their level of confidence and faith in the system increases in tandem (Lee-Archer 2013). The current study was, thus, fittingly focused on evaluating customer satisfaction in mobile and internet service providers.

2.3 Evaluating customer satisfaction - Zimbabwe case studies.

According to (Biza, 2019), who conducted research into the factors that influence customer loyalty in the telecommunications industry, using Telone Bindura as an example. The study's aim was to figure out what factors influence customer loyalty at Telone Pvt Ltd. The effect of pricing, durability, responsiveness, and mean time to repair was investigated in the report (MTTR). In Bindura Town, a questionnaire was sent to all Telone Pvt Ltd clients or subscribers. There was a 100% response time. The survey included multiple standardized questions in which Telone customers were asked to express their opinions on the variables that influence customer loyalty by indicating by answering how much they agreed or disagreed with a given statement. The primary results reviewed that mean time to repair (MTTR) and pricing of the internet were the major factors that affected customer satisfaction in Telone Pvt Ltd. Another study was also undertaken in Zimbabwe by (Mazikana, 2020) to assess the impact of service quality and customer satisfaction on consumer purchasing decisions in Zimbabwe telecommunications. One of the objectives was to establish the steps that have been taken by telecommunications operators to increase their customer satisfaction indices. According to (Mazikana, 2020) The target population of this study was 150 employees and a sample of 80 was selected using simple random sampling. Self-administered questionnaires were used as research instruments to collect data from Econet Wireless Zimbabwe, Telecel and Netone Zimbabwe. A total of 80 questionnaires were distributed and 75 of them were valid which translated to 94% response rate. The study revealed that customers had high expectations on empathy, responsiveness and assurance service dimensions. Based on p-levels alone, Chitura, Dube, and Chari (2007) discovered that network efficiency has the greatest effect on both service quality and consumer satisfaction of the three outcome/technical quality dimensions defined in the qualitative stage. In addition, the study discovered a connection between cell phone efficiency and customer satisfaction. Finally, the results show that by using a service provided by their network operators, cell phone users face a slew of network-related issues. This means that mobile network providers are failing to provide customers with the amount of service efficiency that they expect. Hence, subscribers' service quality perceptions and their satisfaction levels are low. Due to a variety of switching hurdles, customers have not migrated from their existing network providers. As a result, it is fair to conclude that cell phone users are prisoners, stuck in an unfavourable situation. Another research [54] (Makanyeza & Chikazhe, 2017) investigated the mediators of the impact of service quality on customer satisfaction among Zimbabwean bank customers.

Service efficiency, retention, and corporate reputation all have favourable direct impacts on loyalty, according to the report. The impact of service quality on loyalty is also mediated by satisfaction and corporate image, according to the findings. All studies above were carried out to explore how various organisations evaluate their customer experience journeys, their loyalty to the brand and the satisfaction they have on the service quality. Having such knowledge will enable service providers to prioritise on customer retention initiatives with business strategies that aim to improve customer satisfaction in the process improve profitability. Another research (Hwambo, Shamhuyenhanzva, & Sandada, 2017) looked into the impact of consumer frustration, low switching costs, lack of customer service, lack of appropriate or adequate ads, and increased security/ethical issues on customer churn in Zimbabwe's mobile telecommunications market. Structured questionnaires were used to gather data from a group of 413 Zimbabwean mobile subscribers. Customer frustration and a weak complaints management mechanism were found to be important determinants of customer churn. According to the report, mobile service providers should prioritize programs to improve consumer loyalty and invest in processes and approaches that efficiently and satisfactorily address customer grievances, ensuring customer retention. Another research (Jena, 2017) used a case study of MBCA bank to assess the contribution of ICT to customer loyalty in the banking sector in Zimbabwe.

The study was able to demonstrate that information and communication technology (ICT) is one of the variables that lead to customer loyalty in the banking industry. The study recommended MBCA bank to understand its customers through research, building a customer relationship and act on the expectations of customers rather than of the bank itself and further recommended the study into the evaluation of other factors that also affect customer satisfaction in the banking sector. Another research (Gwambuka, 2017) looked at the impact of internet banking on customer loyalty at the ZABG bank in Harare. The study's goals is to figure out how much internet banking contributes to consumer satisfaction and what the relationship is between internet banking and customer satisfaction. Another analysis conducted in Zimbabwe on behalf of Barclays Bank (Nyamukondiwa, 2012) looked at the effect of customer relationship marketing on customer loyalty. The study used a sample 15000 with sample size of 375 for customers. The sample for employees was 160 with a sample size of 113. The findings showed that customer relationship marketing had a positive impact on customer satisfaction. In again another study, (Viriri & Phiri, 2017) investigated the determinants of customer satisfaction in the Zimbabwean telecommunication industry. The study was done on the basis that due to the impact of competition, deregulation, privatization and globalization; customer loyalty is vital in highly competitive industries like the telecommunication sector. The population of the study included six telecommunication firms in Zimbabwe. Data was gathered from telecom customers, staff and senior management. The findings of the study revealed that most subscribers preferred reduced tariffs, sales promotions, improved network coverage and service delivery. The research concluded that most telecom customers are dissatisfied with the services being offered by firms.

All the case studies cited above attempted to evaluate the effects of various service metrics such as service quality on customer satisfaction. The studies were carried out from a service provider's perspective. This means the studies themselves were already biased towards what the service providers thought affected customer satisfaction. Evaluating customer satisfaction using sentiment analysis brings out multiple perspectives as a result of the encounters that take place between the consumer and the service providers .The data supplied by customers on social media is raw and unfiltered. There is a higher chance of service providers improving customer satisfaction as these affords them to capture issues affecting their customers in real time and respond accordingly.

2.4 Evaluating customer satisfaction using sentiment analysis - International case studies.

Consumer satisfaction is an indicator of how well a company's goods and/or services match or exceed customer expectations. (Al-Otaibi et al., 2018) conducted a study in which he used sentiment analysis to assess consumer loyalty. A research related to this was also conducted by (Anastasia & Budi, 2016). The researchers wanted to use sentiment analysis of Twitter data to gauge customer loyalty.The businesses in question used Twitter to communicate with their clients and advertise their services. Between February and March 2016, 126,405 tweets were collected. The Net Sentiment Score was determined. The experiments showed that Grab's customer satisfaction is higher than GO-JEK's. Customers prefer to list both the company's Twitter account for negative experiences and not the company's account for good feedback, according to the report.

Similarly, (Feine, Morana, & Gnewuch, 2019) conducted a study using Sentiment Analysis to measure service encounter satisfaction with customer service chatbots.

Most businesses use chatbots to interact with their clients.The key goal of this study was to provide an alternative solution to service providers that are having difficulty assessing chatbot service encounter satisfaction (CSES), as most methods are restricted to post-interaction surveys that are seldom completed and often biased. Chatbots make it difficult for service providers to react rapidly to service problems and unhappy customers.

The paper contributed by proposing and implementing an automated and impartial method for measuring CSES using sentiment scores as a proxy. Customer reviews are recognized as fruitful information sources for tracking and improving customer satisfaction levels, especially because they convey the real voices of individual consumers voicing reasonably unambiguous opinions, according to a related study on customer satisfaction (Kang & Park, 2014). Sentiment analysis has risen to prominence as a statistical tool for analysing consumer reviews. While multiple sentiment analysis methods have suggested extracting emotional insights from consumer feedback, the question of how to successfully analyze customer reviews for the purpose of monitoring customer loyalty with mobile services remains unanswered. As a result, their research established a new method for assessing customer loyalty based on customer feedback of mobile app providers. They argued that a customer-review-driven solution saves time and money in assessing customer loyalty while also capturing the actual voices of consumers, based on the findings of their research.Another study on measuring customer satisfaction was done by (Gitto & Mancuso, 2017) where they wanted to improve airport services using sentiment analysis. The basis of the study was on the realisation by researchers that measuring the level of customer satisfaction of airport passengers provides a valuable feedback to airport managers and recognized that measuring airport performance through purely operational approaches is not sufficient. In a separate study (Miranda & Sassi, 2014), the researchers used Sentiment Analysis to suggest a method to help in the assessment of customer loyalty in a Brazilian online job search business. The study's foundation was built on the fact that the internet has facilitated the rise of many online service businesses. The researchers looked at client feedback received from a service cancellation form in a database of an online work search firm. This database included a score assigned by the customer as well as a note about the facilities, among other things. They used Python-based tools to classify the sentiment reflected in user feedback, and then measured the sentiment score's association with the clients' scores. The findings led to the suggestion that using sentiment analysis as a support instrument to enrich customer loyalty assessments was beneficial.

Aside from conventional techniques of consumer targeting, social media offers its own range of possibilities. Consumers express their views on utilities and goods in general and with their friends and family. This useful information can be used to help make business decisions. However, it is difficult to retrieve useful information from large volumes of unstructured data. Rather than simply reading and counting email, Social Media Analytics extracts information from social media data and analyses emotion.

2.5 Internet connectivity and demand - Teleworking

COVID-19 cases are on the rise and business operations for mobile and fixed internet providers in Zimbabwe has been significantly affected by government enforced lockdowns. Most businesses have been forced to adopt teleworking. Teleworking has its challenges, most businesses now depend on mobile and fixed internet providers such as Econet, ZOL and Telone for reliable communications services. This has caused a huge strain on their customer services team as there are trying to serve every customer scattered across Zimbabwe. However, this is the best time for communications providers as more subscribers both individual and corporate are looking for reliable service providers that offer the best customer services. The battle for customers started a few years back when POTRAZ requested monthly metrics regarding service quality. This has led providers to revamp their social media-based market research and customer experience programmes. Sources of data that can be used for evaluating customer experience are now being augmented by social media data. Most organisations have now embraced social media because of the ability to influence brand awareness, increased customer loyalty, positive brand association, positive perceived quality and other competitive advantages. Most of the previous research done on evaluation of customer satisfaction in Zimbabwe, service quality and cost metrics were found negatively influence customer satisfaction.

2.6 How service quality affects customer satisfaction

According to Karim and Chowdhury (2014), service quality is the product of a consumer's contrast of their preferences and perceptions. According to Shahin (2006), service quality refers to how well a service satisfies the demands or desires of consumers. Service quality, according to Shahin & Janatyan (2011), is described as the difference between customers' perceptions of service and their expectations. According to Karim and Chowdhury (2014), a customer's expectation acts as a basis for measuring service efficiency so quality is high when performance exceeds expectation and poor when performance falls short of the customer's expectations. According to Shahin and Janatyan (2011), the first step in improving service quality is study and evaluation, which necessitates the use of instruments to assess service quality. In the social security sector various quality initiatives have been instituted to improve customer satisfaction and guidelines have been developed to improve quality in the sector (Lee-Archer 2013). These efforts have culminated in the launch of the International Social Security Association (ISSA) Centre for Excellence at the 2013 World Forum, which sets service quality standards in social security organizations. (Lee-Archer, 2013). The service quality guidelines put people at the centre of the social security system. The conventional service model, in which the social security organization was at the center, is being replaced by an environment approach, in which individuals are at the center, as Lee-Archer (2013) puts it. People will use the ecosystem's partners to completely engage in the social security framework in this setup. According to Lee-Archer (2013), many social security organizations around the world are investing heavily in service efficiency in order to foster a culture of trust and faith that will lead to the general wellbeing and prosperity of the social security sector.The model appositely recognises the importance of service quality in social security provision. The need for quality service is made more compelling by the fact that contribution to social security schemes is compulsory so contributors naturally expert a good service from NSSA, compulsorily collects their contributions.

2.7 Relationship between Service Quality and Customer Satisfaction

According to Karim and Chowdhury (2014), service efficiency and customer loyalty have long been recognised as critical to growth and longevity in today's dynamic market. They also argue that customer loyalty is dependent on the consistency of the service provided. Customer satisfaction, according to Saravanan and Rao (2007), is determined by the level of service quality provided by service providers. Wilson et al. (2008), quoted in Karim & Chowdhury (2014), argue that consumer loyalty and service efficiency share certain characteristics. They do point out, though, that customer loyalty is a wider term than service efficiency. Again, according to Zeithaml & Bitner (2003), cited by Karim &

Chowdhury (2014), while other factors like price and product quality can influence customer loyalty, perceived service quality is a component of it. Daniel and Berinyuy (2010) also agree that there is a correlation between service quality and customer satisfaction, emphasizing the value of including customer satisfaction when defining service quality. Thus, to measure customer loyalty in this analysis, service efficiency dimensions were used. The measurements of the SERVQUAL model, which was used in this analysis, are tangibles, reliabilities, and consistency, responsiveness, assurance and empathy

2.8 Effect of Service cost on customer satisfaction

Pricing strategy is one of the most important aspects of a business. In some ways, we are rational when it comes to making a purchase decision. We are not the same in other respects. When two products are placed next to each other, most consumers would choose the more expensive one, even though it is not. According to research, as costs rise, consumers' perceptions of the quality of the goods on offer rise as well. Customers will be more mindful of the product's overall quality if you have very low prices, and they will be more likely to spot flaws or possible shortfalls. Customer loyalty has long been regarded as a key indicator of a company's performance, and it is generally accepted that it contributes to higher and more reliable sales.

2.9 The SERVQUAL Model of Customer Satisfaction

SERVQUAL was created in the 1980s by Parasuraman, Berry, and Zeithaml, according to Gibson (2009). SERVQUAL is a multi-item scale designed to measure consumer expectations of service efficiency in service and retail establishments. The SERVQUAL model originally consisted of ten service quality dimensions: tangibles, dependability, accessibility, connectivity, reputation, protection, integrity, courtesy, customer comprehension, and access (Daniel & Berinyuy, 2010). Since certain dimensions overlapped, these dimensions were later reduced to five (Daniel & Berinyuy, 2010).

The below are the five dimensions:

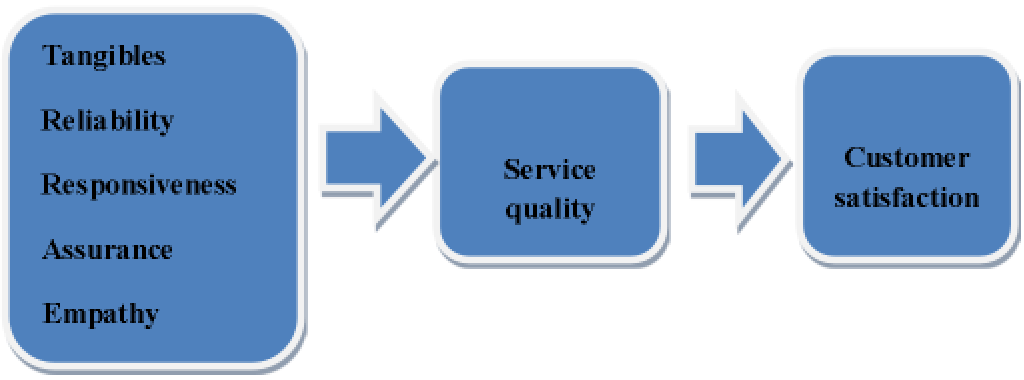
- **Tangibles**, such as physical buildings, appliances, and the presence of employees.
- **Reliability** - The ability to provide consistent and accurate service.
- **Responsiveness** - Willingness to assist and respond to the needs of customers.
- **Assurance** - The ability of employees to instil faith and trust in others.
- **Empathy** - The degree of which a compassionate, one-on-one treatment is provided.

The SERVQUAL technique, which is used by many organizations, is the most prevalent tool for calculating service quality, according to Shahin (2006). SERVQUAL, according to Shahin (2006), is a standardized instrument with reasonable reliability and validity and a wide range of applications. Its aim is to act as a diagnostic tool for identifying large areas of a company's service quality weaknesses and strengths. SERVQUAL dimensions and products reflect key measurement metrics that transcend individual businesses and sectors, according to Shahin and Janatyan (2011, p.101), and have thus been used to assess service efficiency in a wide range of service environments.

2.9.1 Conceptual Framework

As part of the conceptual framework the five dimensions of quality identified by Parasuraman, Berry & Zeithaml in the 1980s under the SERVQUAL model were used to assess service quality and customer satisfaction. The dimensions are as follows: tangibles, reliability, responsiveness, assurance and empathy.

Figure - The SERVQUAL model is the conceptual framework



Hoyer and MacInnis (2001) said that satisfaction can be associated with feelings of acceptance, happiness, relief, excitement, and delight. The servqual conceptual framework above can be used to complement sentiment analysis by applying the model to sentiment analysis results so as to classify which dimension each customer sentiment falls in.

2.10 Social Media Analytics

According to Guy (2012), social media is a series of Internet platforms, programs, and activities that promote teamwork, group development, engagement, and networking. It is originated from the social software revolution. In a similar vein, Kaplan and Haenlein (2010) described it as "a category of internet-based applications that enable the development and sharing of user-generated content." It encourages internet users to communicate with one another. Kapoulas and Mitic (2012) went on to say that what started as a collection of channels for online interactions with an emphasis on entertainment quickly grew into a global phenomenon in which being linked to online networks is something and the ability to "follow," "like," or "post" means influence. This alone indicates that social media has grown in importance around the world. Mayes (2011) emphasized this increase of social media by stating that it is constantly evolving and that new platforms are being created on a regular basis. According to Mayes (2011), this has made it possible for businesses to not only be comfortable with certain sites, but also to understand how to reach out to consumers through them. People spend more than a third of their waking day viewing content on social media, according to Laroche et al. (2013). According to Levinson and Perry (2011), social media is now referred to as "online media" in order to distinguish it from the other two forms of media. TV, journals, and magazines are examples of ‘‘old media," according to Levinson and Perry (2011). This forms of media are created and controlled by experts, emphasizing the top-down management approach. Email, blogs, internet message boards, chat rooms, and other forms of ‘‘new media" are examples of the second kind. ‘‘New media" transcends the time and place of ‘‘old media" as a product of internet technologies (Zhou and Wang, 2014). According to Ruddell and Jones (2013), ‘‘new media" includes blogs and microblogs (e.g., Twitter), Wikis (e.g., Wikipedia), photo and video sharing websites (e.g., Flickr), BBS (e.g., Tianya in China), social networking sites (e.g., Facebook), and internet groups (e.g., Maopu in China). Despite its increasing prevalence and widespread acceptance of social media's effects, organizations and customer service administrators have yet to develop a formal understanding of how to use social media to measure customer engagement.

2.11 Social Media Analytics and Sentiment Analysis.

There is often an implicit feeling as consumers interact with businesses by leaving suggestions. You may use emotion analysis to collect this intangible data to better understand the clients. Why are they dissatisfied with customer service or with the product in general? Whatever aspects of the product do they prefer? Customer retention and engagement can be improved by understanding and acting on the insights presented by sentiment **analysis**. Sentiment analysis can help track a customer sentiment on social media in real time which can help an organisation instantly spot a critical issue affecting customers and, in the process, take immediate action. Most organisations in zimbabwe have already embraced social media when engaging their customers. Social media offers raw and unfiltered customer feedback as compared to data gathered from call logs or net promoter score (NPS) surveys. According to (Stelzner, 2013), social networks have become an integral part of daily life, serving as a platform for the exchange of news, thoughts, and knowledge of all sorts.

2.12 Sentiment analysis techniques

Most of the literature on sentiment analysis relates to developed countries, there is no literature that relates to the use of sentiment analysis in evaluating customer satisfaction in Zimbabwe. This research aims to fill this gap in literature through carrying out a case study on the effectiveness of evaluating customer experience using sentiment analysis in mobile and internet service providers. This study was an investigation on the effectiveness of social media on how it can be used to measure customer satisfaction and enhance customer experiences in this industry. Sentiment Analysis also known as opinion mining has become a topic of interest as an emerging field that has many practical applications. It explores people's sentiments, opinions, behaviours, attitudes, and emotions towards individuals, organizations, products and services. The fundamental objective of this research is to extract opinions of users on twitter analysing the data for user sentiments and eventually create a model that will be used to evaluate and measure customer satisfaction by telecommunications providers in zimbabwe. When applied to customer experience, the resultant model aims to provide deep insights into how customers perceive products and services which otherwise the same cannot be established using traditional methods. Traditional approaches such as analysis of call logs or conducting customer surveys are subject to researcher presence and small sample sizes who are self-aware. Sentiment analysis addresses these problems by systematically collecting and analysing online sentiments as emotions expressed on that product or service. Since the participants are self-aware, there is a distortion in the data collection, while sentiment analysis provides an accurate and reliable assessment of customer opinions in real time, allowing advertisers to collect insight on perceptions and opinions as they happen without having to engage in time-consuming and expensive market polling (Rambocas, 2013).Opinion mining, according to (Perumal, 2010), is a subfield of Natural Language Processing (NLP) that deals with analysing text data to determine polarity and includes sentiment analysis. Analysis of sentiment is widely used in mining of subjective information from internet content using various techniques including NLP, Statistical techniques and Machine Learning methods. Any opinion/review given by any of an individual through which the feelings, text message, attitudes and thoughts can be expressed is known as sentiment (Tyagi, Chakraborty, Tripathi, Choudhury, 2019). Sentiment analysis has grown in popularity due to the proliferation of microblogging sites such as twitter. Businesses have realised value in sentiment analysis as it provides insights that speak to the inner views of the audience. The results enable businesses to tailor their services or products based on their clientele preferences. Sentiments can be categorized into three categories, positive, negative and neutral. Consumer interaction happens in real time via social media. This form of contact provides a once-in-a-lifetime marketing intelligence tool (Rambocas, 2013). Customer emotions are indirect motivators of purchasing behaviour for the marketer, and the role of emotions in market analysis is not recent. Emotions and powerful brands (Aaker and Keller, 1990; Morrison and Crane, 2007); emotions and consumption; and emotions on product reviews have all been identified (Mano and Oliver, 1993). Consumer sentiment is of value for marketers gathering market intelligence but greater value for consumers who need to gain a quick overview of the collective opinion about a provider or product/service/experience. Due to the vast amount of consumer-generated content available, extracting sentiment manually has become impractical, encouraging research into the applications of sentiment analysis.

In academia and practice, there is a growing interest in and importance of sentiment analysis, according to academic literature (Leung, Law, Van Hoof, & Buhalis, 2013). There is also a growing number of software companies that are offering sentiment analysis tools (Liu, 2010). While tremendous progress has been made in sentiment analysis due to advances in natural language processing (NLP) , current approaches are still far from reaching human-like abilities to discern specific sentiments. And even humans are not that great at determining sentiment, as it is inherently subjective (Tsvetovat, Kazil, & Kouznetsov, 2012). As a result, opinion analysis continues to be a hotbed in academic and industrial research and development. Many businesses are beginning to recognize how customer voices influence other customers' purchasing decisions. In the consumer industry, sentiment analysis is used for product ratings, advertisement patterns, and social media.

According to (D'Andrea, Ferri , Grifoni , Guzzo, 2015) , there are three types of techniques for sentiment classification:

2.12.1 Machine learning approach,

2.12.2 Lexicon based approach and

2.12.3 Hybrid approach.

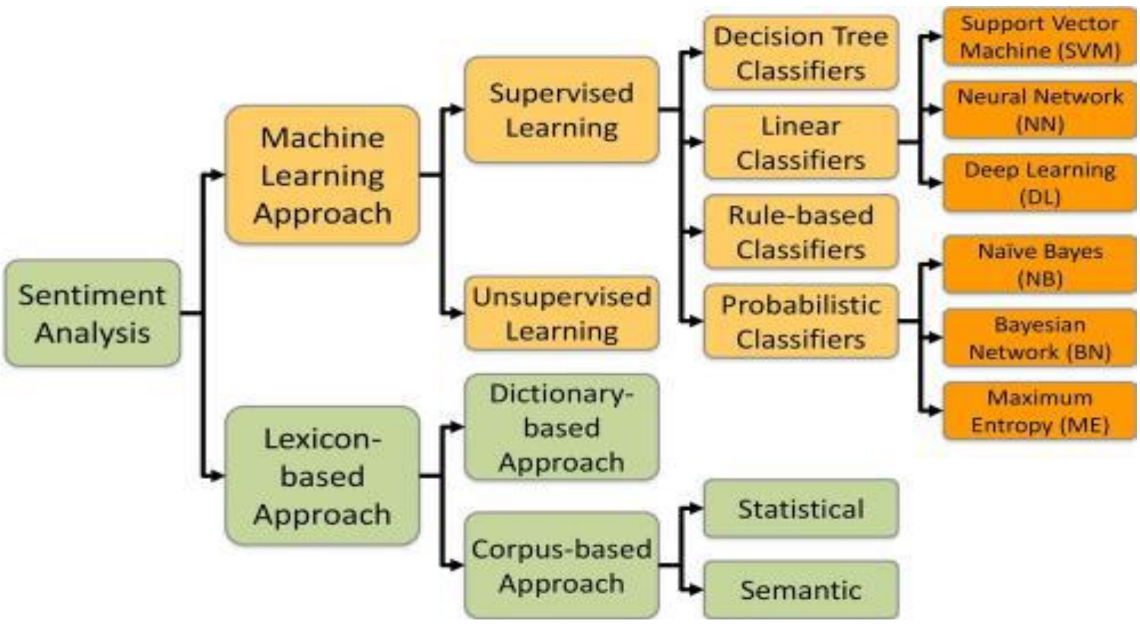


Figure 1: Sentiment Classification techniques (Medhat et al.2014).

2.12.1 Machine learning approach

Centred on both qualified and test data sets, a machine learning technique is used to simulate the polarity of sentiments. It employs machine learning algorithms as well as linguistic functions. The ability to adjust and build trained models for unique uses and situations is the method's key benefit, while its main drawback is the method's limited applicability on new data due to its reliance on the supply of marked data, which can be expensive or even prohibitive. It may use both supervised and unsupervised techniques. When there is a finite number of groups, machine learning uses a supervised approach (positive and negative). To train classifiers, this method requires labelled data. A training set is used by an automatic classifier to learn the various characteristics of papers, and a test set is used to verify the output of the automatic classifier in a machine learning based classification. When finding labelled training manuals is impossible, unsupervised approaches are used. To mine the results, unsupervised learning does not require previous instruction. The semantic orientation (SO) of individual phrases inside the text is determined in unsupervised approaches to document-level sentiment analysis.
The text is classified as positive if the average SO of these phrases exceeds a predetermined threshold, otherwise it is classified as negative. Among the machine learning approaches the most used are:

(2.12.1.1) **Bayesian Networks**: it is a probabilistic approach that models' relationships between features in a very general way. It is based on directed acyclic graph in which nodes are variables and arcs represent the dependence between variables.

(2.12.1.2) **Naive Bayes Classification**: it is an approach particularly suited when the dimensionality of the inputs is high. Despite its simplicity, it can often outperform more sophisticated classification methods.

(2.12.1.3) **Maximum Entropy**: this method is mostly used as alternatives to Naive Bayes classifiers because it does not assume statistical independence of the random variables (features) that serve as predictors. The principle behind Maximum Entropy is to find the best probability distribution among prior test data.

(2.12.1.4) **Neural Networks**: this model is based on a collection of natural/artificial neurons uses for mathematical and computational model analysis

(2.12.1.5) **Support Vector Machine:** SVM is a supervised machine learning algorithm that can be used for both classification or regression challenges. Classification is predicting a label/group and Regression is predicting a continuous value. SVM performs classification by finding the hyper-plane that differentiate the classes we plotted in n-dimensional space. It finds an optimal solution.

Table 6 - Comparison of Different Machine Learning Methods

Methods	Advantages	Disadvantages
Support Vector Machine Method	High-dimensional input space. • Few irrelevant features. • Document vectors are sparse	A large amount of training set is required. • Data collection is tedious
N gram SA	Usage of 1- and 2-grams as features for sentiment prediction can increase the accuracy of the model in comparison with only single word feature.	Long range dependencies are not captured. • Dependent on having a corpus of data to train from.
Naïve Bayes Method	Simple and intuitive method. • It combines efficiency with reasonable accuracy.	Mainly used when the size of the training set is less. • It assumes conditional independence among the linguistic features.
Maximum Entropy Classifier	This method do not assume the independent features like NB method. • Can handle large amount of data.	Simplicity is hard.
KNN Method	Based on the fact that the classification of an instance will be somewhat similar to those nearby it in the vector space. • It is considered computationally efficient.	Large storage required. • Computationally intensive recall.
Multilingual SA	The texts of different languages are evaluated without translation.	Training corpus for different language is needed.

	• Deals with 15 different languages.	
Feature Driven SA	Adaptable to large projects. • It is a concise process.	Not a powerful on smaller projects.

2.12.2 Lexicon based approach.

The lexicon-based methodology, on the other hand, does not require any previous preparation to mine the details. It works from a pre-defined list of terms, each of which is linked to a particular emotion. They work by counting the number of positive and negative terms in a sentence. These techniques differ depending on the context in which they were created. While lexical does not need marked details, it is difficult to produce a specific lexical-based dictionary that can be used in a variety of contexts. Slang used in social media, for example, is scarcely assisted by lexical processes. For deciding polarity, the lexicon-based approach uses a sentiment dictionary of opinion terms and matches them to the results. Manual creation, corpus-based methods, and dictionary-based methods are the three methods for creating a sentiment lexicon. Construction by hand is a complex and time-consuming process. Opinion terms can be produced with a high degree of accuracy using corpus-based methods. Finally, in dictionary-based approaches, the aim is to manually compile a limited collection of opinion terms with defined orientations, then expand this set by looking up synonyms and antonyms in the WordNet dictionary.

2.12.3 Hybrid approach.

Finally, the hybrid approach has the ability to increase emotion classification accuracy by combining machine learning and lexicon-based approaches. Depending on the intent of the study, there are certain benefits and disadvantages to using these various methods. We'll give you a quick rundown of the most important points. The key benefit of machine learning techniques is the ability to customize and develop qualified models for particular purposes and situations, while the main disadvantage is the difficulty of incorporating basic information that is not learned from training data into a classifier. Furthermore, since they often rely on domain specific features from their training data, learned models often have low adaptability between domains or text genres. Lexicon-based approaches have the advantage of covering more terms in general knowledge emotion lexicons; however, these approaches have two major drawbacks. To begin with, lexicons have a finite number of terms, which can pose a challenge when extracting emotion from highly diverse settings. Second, regardless of how words are used in a paragraph, sentiment lexicons prefer to give a set sentiment orientation and ranking to them. The lexicon/learning symbiosis, the recognition and evaluation of emotion at the definition level, and the reduced exposure to changes in the subject domain are the key benefits of hybrid methods. The method's key drawback is that reviews with a lot of noise (words unrelated to the review's subject) are often given a neutral score because the method fails to sense any emotion.

2.13 Comparison and Consolidation

Table 7 shows a comparison and consolidation of the three major approaches to emotion analysis. Using diverse methods to perform sentiment analysis can yield varying results. Each method has its own set of advantages and disadvantages. The machine learning approach produces the best results by taking into account main variables such as success, reliability, and precision, and it is where the majority of the work has been performed. This activity has developed into many processes, which are listed below.

Table 7 - Comparison of the three Approaches

Approaches	Classification	Advantages	Disadvantages
Machine Learning Approach	Supervised and Unsupervised learning	• Dictionary is not necessary. • Demonstrate the high accuracy of • classification.	Classifier trained on the texts in one domain in most cases does not work with other domains.

Rule Based Approach	Supervised and Unsupervised learning	• Performance accuracy of 91% at the review level and 86% at the sentence level. • Sentence level sentiment classification performs better than the word level.	Efficiency and accuracy depend on the defining rules.
Lexicon Based Approach	Unsupervised learning.	Labelled data and the procedure of learning is not required	Requires powerful linguistic resources which are not always available.

2.14 Chapter summary

This chapter seeks to provide a literature review on sentiment analysis techniques and how its applied in the field of customer analytics. It also aims at highlighting some of the challenges of the different sentiment analysis techniques that remain in advancing current approaches. Work done by several researchers in the past was reviewed with regards to customer satisfaction and sentiment analysis. It gave the researcher room to evaluate this research in terms applicability in other domains. Past research was evaluated in terms of weakness and strength in the current environment.

Chapter 3 Research methodology

3.0 Introduction

This section explains and justifies methods the researcher used to extract and analyse data for this study. The section focused on the research philosophy, research design, research instruments, research population, sampling, validity and reliability, data analysis, ethical considerations undertaken throughout the research.

3.1 Research Philosophy and Paradigm

This research followed the pragmatism approach because of its potential to allow the mixing of methods given the nature of the research problem that was being investigated. This case study outlined the methods used to both quantitatively analyse Twitter data for overall public sentiment and qualitatively analyse the same data to discover details in the changing discourse when customers are engaged in a certain way. It was necessary to understand the content posted on social media and adopt comprehensive analytical techniques to predict and evaluate relevant information within the same context it was posted. One of the difficulties in analysing textual data is that it often necessitates binary categorization: does it belong in this or that category?

This has always been the greatest challenge for many sentiment analysis researchers, particularly when dealing with data that requires a little more detail to understand.

Both quantitative and qualitative methods were used to analyse data using sentiment analysis using lexical signifiers of emotion to decide whether the data tends towards positive or negative emotions. Qualitative analysis was applied on text extracted from twitter where the researcher had to determine the subjectivity (our emotions) and explaining it using some abstract representations. Qualitative was used to detect semantic patterns and context of user tweets while quantitative was used to determine the percentage of negative and positive tweets.

3.2 Research design

The researcher adopted the descriptive observational research design where tweets or customer behaviour on twitter were closely observed without influencing them in any way. Customers on Twitter share information about their preferences or opinions on products and services voluntarily. Thus, they are honest in their views which fits well into the characteristics of this design.

While there are 3 methods that have been used by various researchers regarding sentiment analysis as explained in chapter 2, the researcher adopted the lexicon-based approach, the most primitive of the 3 methods. The main objective of the study sought to justify why mobile and fixed internet service providers need to adopt sentiment analysis in evaluating customer experience given the increase in reported tickets and not necessarily a comparative study of the efficiency of the 3 methods, hence the decision to stick to the lexicon-based method.

3.3 Sample size and sampling method

The amount of data streaming from twitter is defined as big data which is complex to analyse due to its volume, variety and velocity. Streaming data is data that is continuously generated and has no discrete beginning or end. It varies from time to time. Scalability, continuous availability, diversity, data security and manageability are the challenges in streaming data. Rather than analysing the entire streaming dataset, sampling provides an alternate solution to analyse in an efficient manner and thereby minimizing the computation time. The Reservoir sampling (RS) technique was used to extract batch samples of size 200 per page refresh.

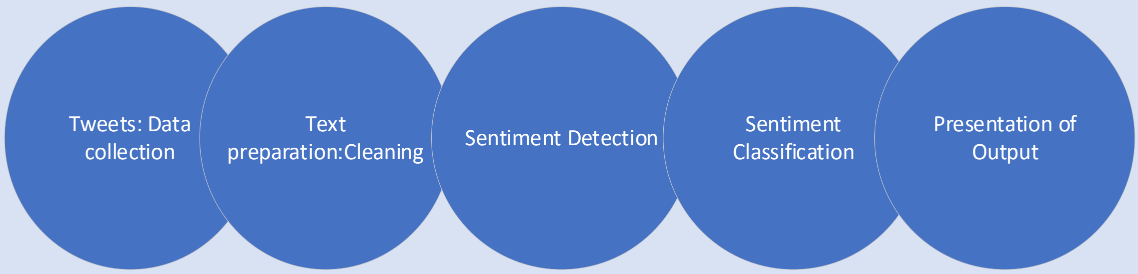
3.4 Research instruments

The Twitter API and the Tweepy python library was used to extract data from Twitter and stored in memory for further processing using python and data analytics libraries such as pandas and NumPy.

Table 8 - Research instruments

Python Library	**Description of libraries**
textblob	Textblob is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.
NLTK	The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for applying in statistical natural language processing (NLP). It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning.
NumPy	This library is going to be used for calculations in multidimensional space.
Pandas	Library used for data manipulation and analysis
Matplotlib	Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

Figure 16 - Sentiment Analysis Pipeline



3.5 Data collection procedures

Data was collected through the official Twitter Application Program Interface (API) using python and then imported into a Jupyter notebook for data cleaning and analysis.

Twitter provides a streaming API that developers can use to download data about tweets in real-time. The researcher obtained Twitter API credentials (API key, API secret key, Access token, and Access token secret) and used the Tweepy library to connect to the Twitter API in real time. Looking at the service provider timelines , there were tweeter engagements between customers and service providers. The researcher chose to store collected data in memory , running analysis as the data was streaming rather than converting the data and storing it into disk. The choice of data collection method was based on the premise that the resultant model will be run on continuous streaming data in production providing analysis that can be used in real time. However as highlighted on the delimitations of this study the researcher chose not to include data streaming mechanism as part of this study as that would have added complexity.

Mobile and fixed internet service providers such as Econet Zimbabwe , Netone Pvt Ltd ,ZOL zimbabwe and Telone Pvt Ltd all have an online presence on Twitter as identified by their twitter handles below.

Figure 17 - Econet twitter handle data extraction

```
from IPython.display import display
screen_name="econet_support"
posts = tweepy.Cursor(api.user_timeline,screen_name, count = 200, lang ="en", tweet_mode="extended" ,include_rts=False)

for tweet in (posts.items(5)):
    display(tweet.full_text)
```

```
'Hey @sircarta_. Please check your other post for our response. ^SC'

'Hey @sircarta_ @nico_lasi. We want you to have the connection you need. Do you mind sharing your mobile number, loca
tion, and the type of device you are using, so that we may assist? We look forward to hearing from you. ^SC'

'Hey @thatgirlRieRie. Please check your DM for our response. ^SC'

'Hey @thatgirlRieRie. Please check your other tweet for our response. ^SC'

'Hey @thatgirlRieRie. Please check your DM for our response. ^SC'
```

Figure 18 - Telone twitter handle data extraction

```
from IPython.display import display
screen_name="TelOneZW"
posts = tweepy.Cursor(api.user_timeline,screen_name, count = 200, lang ="en", tweet_mode="extended" ,include_rts=False)

for tweet in (posts.items(5)):
    display(tweet.full_text)
```

```
'@Muku98706294 Good evening, may we kindly ask you to take note of our inbox conversation. ^MKM'

'@tendai210693 Good evening, we have assisted you in our DM. ^MKM'

'@Spencer21871124 Good evening, may we kindly ask you to inbox us so that we can assist you. ^MKM'

'@Muku98706294 Good evening, we have responded to your DM.^KM\n#MaskUp #StaySafe'

'@AllenSedze Good evening, may you kindly check your DM.^KM'
```

Figure 19 – Zimbabwe Online twitter handle data extraction

```python
from IPython.display import display
screen_name="@ZOLconnect"
posts = tweepy.Cursor(api.user_timeline,screen_name, count = 200, lang ="en", tweet_mode="extended" ,include_rts=False)

for tweet in (posts.items(5)):
    display(tweet.full_text)
```

```
'You can make use of the following link https://t.co/P3hzzM5dHb\n@Potraz_zw'

'Celebrating International Girls in ICT Day\nJoin us Live on our ZOL Facebook page @ZOLconnect  for a webinar with a
topic: Women in the digital World - Challenges &amp; Opportunities https://t.co/rrJxfaUYBj'

'@mylestonzw Good Afternoon. Thank you for contacting us.\nPlease, may you advise what your ZOL Customer ID is so we
can fully assist?^NC'

'Are you a startup or an SME  just getting started &amp; you want to represent your business with a customized email
address and your domain name?\n\nFor example picasso@curtiscreatives.co.zw or https://t.co/Zu09CHc9Td\n\nClick the fo
llowing link to sign up https://t.co/oavetO1X17 https://t.co/80PovHpmdp'

'@HopeNashie Good Evening, please note that your payment has been successfully credited into your account.^NC'
```

Figure 20 - Netone twitter handle data extraction

```python
from IPython.display import display
screen_name="NetOneCellular"
posts = tweepy.Cursor(api.user_timeline,screen_name, count = 200, lang ="en", tweet_mode="extended" ,include_rts=False)

for tweet in (posts.items(5)):
    display(tweet.full_text)
```

```
'@peacemubs Hi @peacemubs ! May you provide us with: Mobile number, name, surname, ID number,  DOB, and current OneMo
ney balance for assistance. ^MM'

'@babaShinso Hi there and thanks for getting in touch. Kindly check your inbox for the feedback we have provided earl
ier on. Thank you, ^MM'

'@Royal_1Gk Makadii @Royal_1Gk. Tirikuita zvose zvinokwanisika kuti tivabatsire. Mazvita, ^MM'

'✅ Promotion runs from 29 March 2021 - 28 June 2021\n✅ Data voucher is valid for 48hrs\n✅Wallet funding options i
nclude, ZIPIT, money received from another OneMoney account, Bank to Wallet\n#ItsOneTime'

'💡 Fund your OneMoney account with ZWL1000 and receive 500MB Free Data valid for 48hrs!\nHere’s some quick promo inf
o for you to start winning;\n✅ Data voucher is received within 24 hrs of participating\n✅ Each customer is eligibl
e to participate and win once every week https://t.co/sWeXZ3M8cu'
```

The researcher collected data from twitter on the timelines of the 4 service providers as shown above. The data comprised engagements done between service provider customer service agents and their customers. The researcher also worked with the assumption that the samples used in this research would fairly generalise the overall customer satisfaction scores for the service providers even though not all customers are on twitter. The resultant sentiment analysis model can be used on other data such as customer call records , incident management systems or live chat data where customers who are not on twitter could be catered for.

Exclusion criteria included

(1) (1) Used the Twitter API's "language" area to filter out tweets that are not written in English. The researcher only worked with English tweets because concentrating on

one language helps us to abstract away the complexities of sentiment analysis in multiple languages.

(2) Slang words were ignored

3.6 Data cleaning (Text preparation)

In all variables, data cleaning entailed looking for missed values, bugs, and outliers. At this stage, all errors must be resolved, and outliers must be detected. Duplicates and incorrect symbols must be eliminated. Cleaning the extracted data until analysing it was part of the text preparation process. The textual dataset was stripped of any material that was not considered important to the study's subject. Twitter data is often messy and involves a variety of duplicate data that needs to be cleaned before processing in the case of streaming data. Using the panda's collection, non-textual contents and contents that were incidental to the study were found and deleted.

a) Removing Twitter Handles (@user): Twitter handles do not contain any useful information about the nature of the tweet, so they can be removed.

b) Removing Punctuations, Numbers, and Special Characters: The punctuations, numbers and even special characters are removed since they do not contribute to differentiating tweets.

c) Tokenization: Tokenization is the act of breaking down a text paragraph into smaller bits such as phrases or sentences. A token is a single object that serves as the foundation for a sentence or essay. A tokenization outcome is seen in the illustration below.

Input: [their service is poor]
After tokenization: [their service is poor]

d) Stemming: It is a rule-based process of stripping the suffixes (“ing”, “ly”, “es”, “s” etc) from a word.

For example: “play”, “player”, “played”, “plays” and “playing” are the different variations of the word – “play”. The objective of this process is to reduce the total number of unique words in our data without losing a significant amount of information.

Figure 21 - Data cleaning (Text preparation)

```
In [108]: # Create a function to clean the tweets

          def cleanTxt(text):
              text = re.sub('@[A-Za-z0-9]+', '', text) #Removing @mentions
              text = re.sub('#', '', text) # Removing '#' hash tag
              text = re.sub('RT[\s]+', '', text) # Removing RT
              text = re.sub('https?:\/\/\S+', '', text) # Removing hyperlink
              text = text.lower()


              return text

          # Clean the tweets
          df['Tweets'] = df['Tweets'].apply(cleanTxt)

          # Show the cleaned tweets
          df.head()
```

Out[108]:

	Tweets
0	hie _ . _ we have responded. ^nza
1	hie , our apologies for the experience. may yo...
2	hi , we have reversed the transaction back to ...
3	hi , may you please dm us your mobile number a...

3.7 Sentiment detection

The researcher looked at the selected sentences from the reviews and opinions. Sentences containing subjective statements (opinions, values, and views) were kept, while sentences containing factual correspondence (facts, information) were removed. Sentiment identification is performed at various stages, including single terms, words, full sentences, and whole documents, using widely used techniques such as:

- **Unigrams**: Each variable is expressed as a function vector based on the frequency of a single term in this method. It's often referred to as a "pack of words" strategy.

- **N-Grams**: In this approach the features of a document is represented by multiple words in sequence (e.g.: words in pairs, triplets) which captures more context

- **Lemmas**: Rather than using the literal title, synonyms are used. For instance, better -> good and best -> good. This approach, according to the researchers, makes grouping and generalization simpler. Kushal et al. (2003), on the other hand, proposed that meanings are not all synonyms, and used an experiment to show that when terms are added to their thesaurus meanings, emotion classification precision is diminished.

- **Negation**: The phrases "I like this book" and "I do not like this book" may have been considered equivalent in other labelling approaches, but with negation, both words are put into opposing groupings. Negation, on the other hand, is not always simple to model. When sarcasms and ironies are used in a sentence, for example, Pang and Lee (2008) found it difficult to distinguish negation. Furthermore, the negation concept does not necessarily result in polarity reversal. In the statement "No wonder this is considered to be the best book," attaching the word NOT to BEST, for example, would be considered wrong.

- **Opinion words**: There are terms that are used to express people's emotions and viewpoints (nouns, verbs, adjectives, adverbs). The appearance or absence of a word is represented by these words, which are inserted into a function vector. These terms are strong markers of a document's subjectivity. Textual sentences relating to several objects, functions, and attributes are not rare. Sentiment processing can be used to isolate these objects, characteristics, and attributes and categorize them using mathematical algorithms. This aids in the research stages and improves classification and data summarization accuracy.

Figure 22 - Sentiment detection

```
# Create a function to get the subjectivity
def getSubjectivity(text):
   return TextBlob(text).sentiment.subjectivity

# Create a function to get the polarity
def getPolarity(text):
   return  TextBlob(text).sentiment.polarity


# Create two new columns 'Subjectivity' & 'Polarity'
df['Subjectivity'] = df['Tweets'].apply(getSubjectivity)
df['Polarity'] = df['Tweets'].apply(getPolarity)

# Show the new dataframe with columns 'Subjectivity' & 'Polarity'
df
```

	Tweets	Subjectivity	Polarity
0	Hie , our apologies for the experience. May yo...	0.500000	0.000000
1	Hi , we have reversed the transaction back to ...	0.071429	0.000000
2	Hi , may you please DM us your mobile number a...	0.000000	0.000000
3	Hi , our apologies for the experience. Kindly ...	0.641667	0.350000
4	Hi 48697166, kindly provide us with the error ...	0.900000	0.600000
5	Hi 1, our apologies for the late response. Kin...	0.750000	0.150000
6	Hey , may you please let us know your location...	0.000000	0.000000
7	Hey _mhaps @263lod. Kindly note there was no d...	0.725000	0.475000
8	_AK47 Hi 50834056, did you buy the airtime vi...	0.700000	0.300000
9	3, thank you for getting in touch. Please note...	0.450000	0.025000
10	Thank you for your continued support _dhliwayo...	0.758333	0.425000

3.8 Sentiment classification

Subjective sentences are categorized in this process as positive, negative, good, bad; like, dislike; but, classification can be done using multiple points.

Figure 23 - Sentiment classification

Create a function to compute negative (-1), neutral (0) and positive (+1) analysis

```
def getAnalysis(score):
    if score < 0:
      return 'Negative'
    elif score == 0:
      return 'Neutral'
    else:
      return 'Positive'

df['Analysis'] = df['Polarity'].apply(getAnalysis)

# Show the dataframe
df.head()
```

	Tweets	Subjectivity	Polarity	Analysis
0	Hie , our apologies for the experience. May yo...	0.500000	0.00	Neutral
1	Hi , we have reversed the transaction back to ...	0.071429	0.00	Neutral
2	Hi , may you please DM us your mobile number a...	0.000000	0.00	Neutral
3	Hi , our apologies for the experience. Kindly ...	0.641667	0.35	Positive
4	Hi 48697166, kindly provide us with the error ...	0.900000	0.60	Positive

The fourth stage is polarity classification, which divides the textual dataset into classification classes based on each subjective sentence. These classes are usually shown as two extremes of a scale (positive, negative; good, bad; like, dislike). However, various points may be used to classify anything, such as the star scores used for hotels, restaurants, and retailers. In binary and polar classification, several machine learning methods are used. Machine learning is a branch of artificial intelligence that seeks to create predictive models based on previous interactions and observations. It essentially encourages the use of computer programming to study and comprehend the fundamentals of a specific data set, and then to apply that information to forecast or refine a potential criteria. The overall goal is to create a predictive feature that can forecast a desired outcome - y (dependent variable) based on predefined input parameters or attributes - x. (Gama and Carvalho, 2009). This method of learning is known as "supervised learning" where the goal is known. In sentiment analysis, using a supervised leaning technique necessitates a training manual with textual material or a data corpus that acts as a preparation document for classification learning. Naive Bayes (NB), Support Vector Machines (SVM), and Maximum Entropy are the three simple classification functions accessible (ME). A Naive Bayes classifier is a probabilistic classifier that uses Bayes' theorem and assumes that characteristics are independent of the class name. The frequency of occurrence of each attribute per class in the training data set is used to construct this classifier. The mathematical learning theory underpins support vector machines (Vapnik, 1995). By searching for a hyperplane that maximizes the separation margin between observations from various classes, binary classifiers demonstrate strong generalization capabilities. The use of kernels allows for nonlinear problems to be solved. ME creates a set of models, each of which has a function that corresponds to a model constraint. For grouping, the model with the highest entropy of all models is chosen. Since all three classifiers have been validated in the literature (Pang and Lee 2008, Li and Liu 2012), they either require pre-tagged training data or a data corpus, which is not always accessible, or they require a significant amount of time and human capital to construct. Furthermore, the data's language must be considered. The majority of sentiment analysis literature, methods, and techniques are written in English. This creates a challenge when it comes to multilingual translation.

While there is a growing body of research on aligning other languages with the sphere of interest, cross-lingual adaptation remains a challenge, particularly when cultural idiosyncrasies are taken into account (Kim and Hovy, 2006; Blitzer et al. 2007).

The most basic approach is the bag of words method, which assigns a score or weight to each word depending on its nature (good or bad) and occurrence in the text document. After calculating the score for each expression, the arithmetic number or mean is used to determine a score for the whole paper. The most basic scoring approach involves subjectively assigning scores to opinion documents and calculating a "pseudo-expected" rating. Although this approach is scientifically sound and easy to understand, it has been criticized for not being an effective way to categorize vast amounts of data. Furthermore, since it depends on human categorization, the classification's validity has been disputed, considering the diversity of human beings (Li and Liu, 2012). The use of lexicons is another practice. A lexicon serves as a link between a language and the knowledge it expresses. It is a list of words and their definitions in a particular language. For sentiment analysis, a number of lexicons have been created. WordNet is an English-language lexical website. This index, created by Princeton University in 1985, provides general meanings of terms, groups words into synsets of synonyms, and records relationships between synonymy sets using conceptual-semantic and lexical relations.

In 2004, Kamps and Marx applied graph theory to this synonym relationship to determine the distance between terms based on their similarities and distinctions. They first place and adjective on a good-bad (+,-) scale, then calculate the difference between terms using the length of the continuum, with closer words meaning shorter distances. Turney

has launched a new scoring system called Web Search (2002). The semantic issues with single word grouping are recognized by this approach. For example, in a car review, the term "unpredictable" may receive negative feedback, whereas in a movie review, it may receive positive feedback. Turney (2002) used “tuples,” which are adjectives combined with nouns and adverbs combined with verbs, to solve this problem. There are many steps to the word search process. Tuples are first collected from feedback. The semantic orientations of the extracted tuples are then determined, and the average semantic orientations for the whole text are computed. Turney (2002) used the 10 search engine AltaVista and ran two queries to evaluate the semantic orientation of tuples.
One looked at the number of documents that rated the tuple as "excellent," and the other looked at the number of documents that rated the tuple as "bad." It is called a good orientation if the tuple appeared more often in the "excellent" question than in the "bad" query. Similarly, if the tuple appeared more than once in the "bad" question, it would be bad.

3.8 Data presentation and analysis methods

Analysis of data was done to transform the unstructured, scattered twitter text into useful facts. The researcher used pie charts , word cloud , scatter graphs to understand the data. Using python statistical and visualization libraries , data scraped from Twitter was quantitatively analysed using (pandas, matplotlib). Scatter graphs , pie charts, and proportions were used to present the findings.

3.8.1 Dataset

The below dataset sample represents all datasets scrapped from Twitter for Econet , Netone ,Zimbabwe Online and Telone.

Figure 24 - Sample Dataset

	id	myTweets	tweet_date
0	1365589888448479235	hie jamesmaseko kindly note that we have respo...	2021-02-27 09:09:35
1	1365589596344627201	hie jamesmaseko our apologies for the delay pl...	2021-02-27 09:08:25
2	1365587528900870144	hello ajuda the bundles work like an ordinary ...	2021-02-27 09:00:12
3	1365586710344708099	makadini tafarapalesa tinokumbirawo mutarise m...	2021-02-27 08:56:57
4	1365585901452226561	hey morgenmukuli please note that the transact...	2021-02-27 08:53:44
5	1365585645339635713	hie munyamusy we have since responded to your ...	2021-02-27 08:52:43
6	1365585321031835650	hi leisleyb jermainematth muleyaelijah we have...	2021-02-27 08:51:26
7	1365584297466429443	hi vinlyn leisleyb we have reversed the money ...	2021-02-27 08:47:22
8	1365583952673660929	kwazivai prayer tinokumbira mutarise mhinduro ...	2021-02-27 08:46:00
9	1365583358894440448	hi vinlyn we have responded to your dm mbm	2021-02-27 08:43:38

3.8.2 Most common words

Below is the code that was used to extract the most common words from data scrapped from twitter for all 4 service providers. Various data presentation methods such as word cloud , bar charts and line graphs can be used to represent this information. The significance of extracting this information was for the researcher to carry out an analysis to seek out domain specific words which can be used later to generate a domain corpus. People that express themselves in a certain domain often use specific domain language expressions, so creating a classifier that works well across domains is a difficult task.

Figure 25 - Word frequency plot

```
fig, ax = plt.subplots(figsize=(8, 8))

# Plot horizontal bar graph
fd.plot.barh(ax=ax,
              color="purple",
               x='Word',
               y='Frequency',
            )

ax.set_title("Common Words Found (Including All Words)")

plt.show()
```

Figure 26 - Econet word frequency

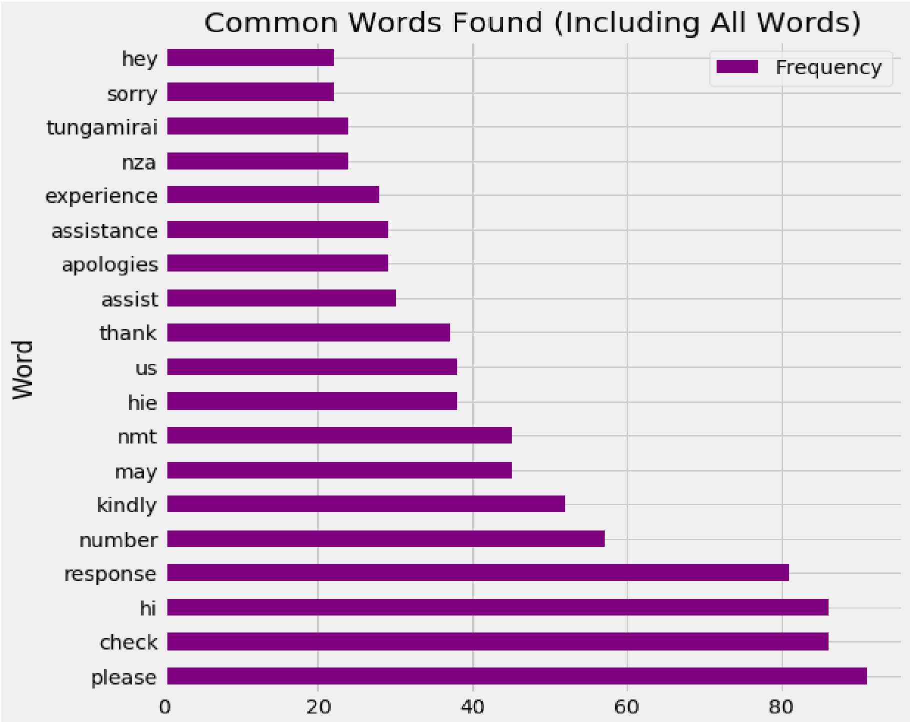


Figure 27 - Netone Zimbabwe word frequency

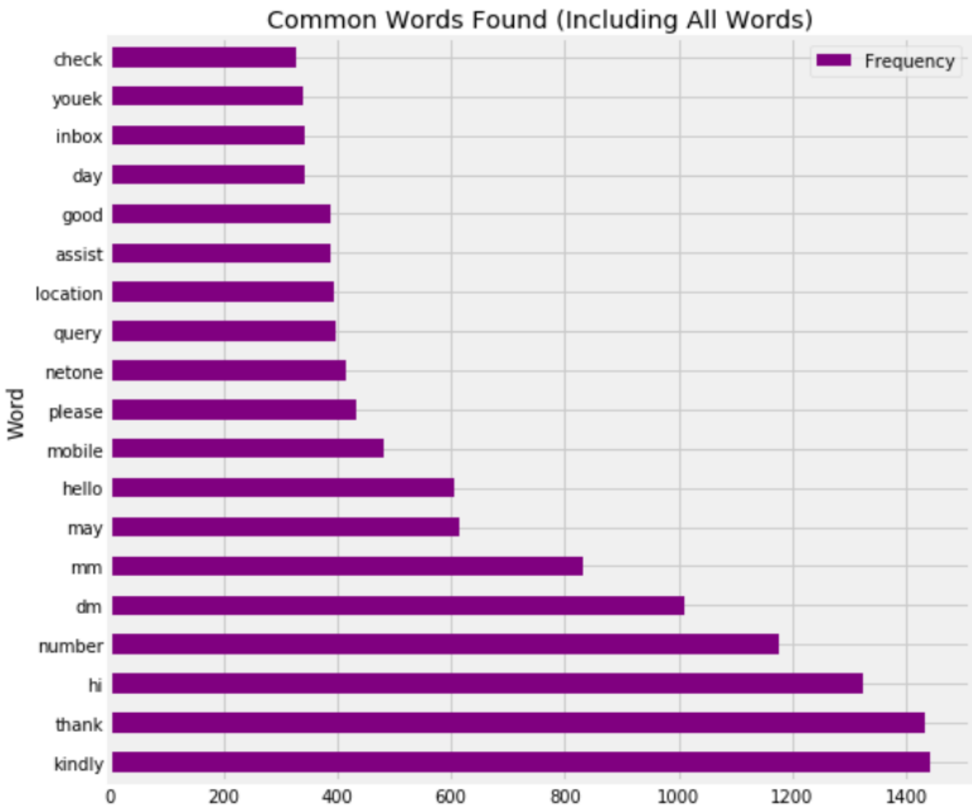


Figure 28 - Zimbabwe Online Word Frequency

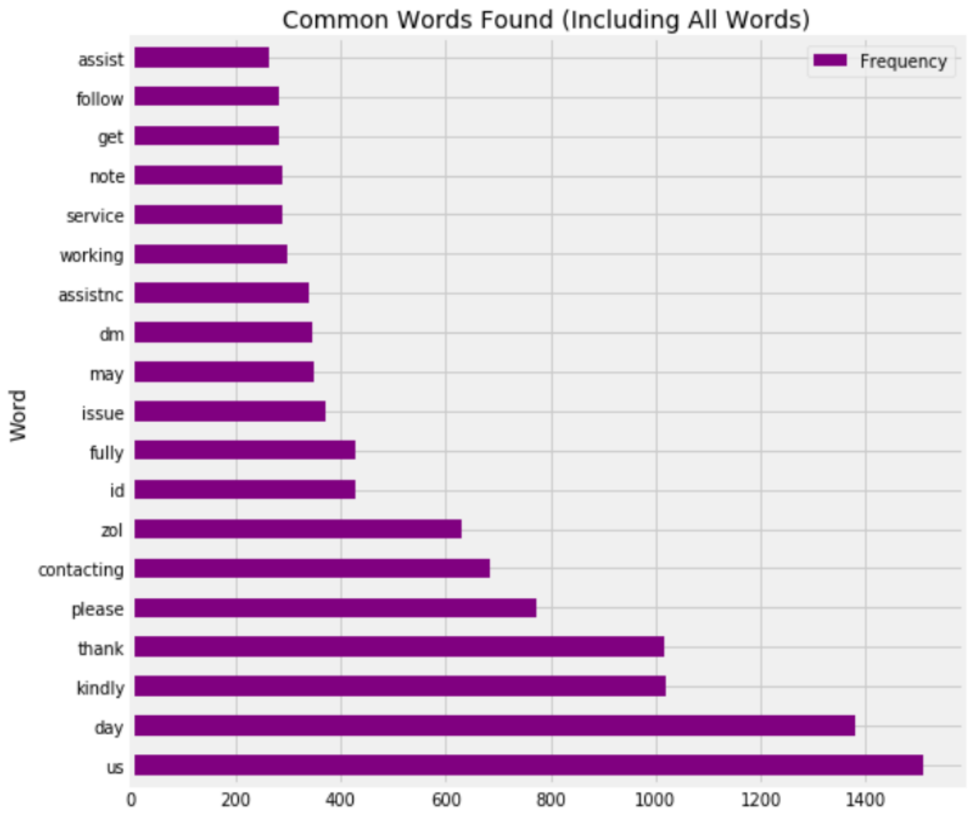
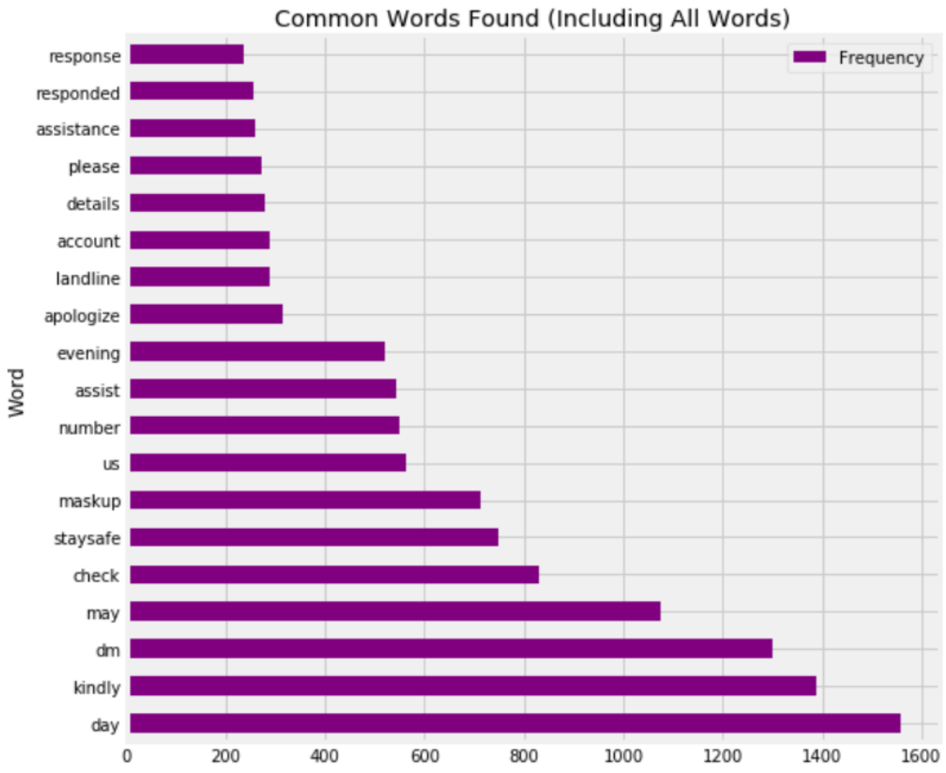


Figure 29 - Telone Zimbabwe word frequency

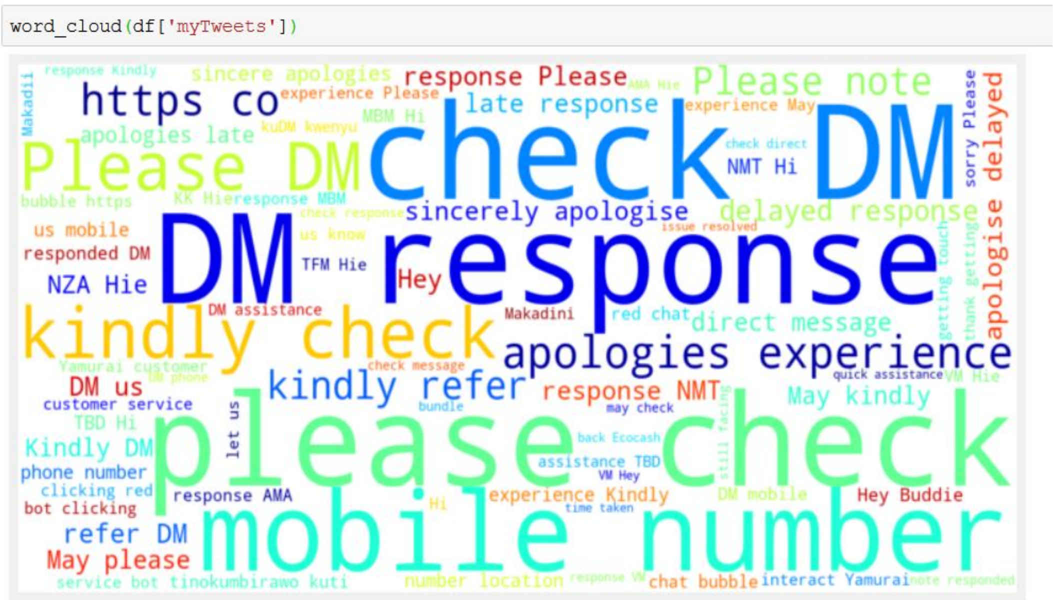


The graphs above for Telone , Econet ,Netone and ZOL show the commonly used words used by customer service representatives when engaging with customers on Twitter. An analysis of these words per provide can help isolate specific services and products that are causing issues to customers and possible find ways of educating the representatives on the words that can be used to improve sentiment.

3.8.3 Word Cloud visualisations

A word cloud (also known as a tag cloud) is a word visualization that ranks the most frequently used words in a text from small to big. The “most famous word(s) graphs above” were used to build the following word cloud. They offer an overview of the most common keywords used by service providers and consumers during engagements in the framework of this study. When comparing two messages, such as service ratings, a word cloud will offer insightful perspectives. Stopwords (a, and ,the, etc.) have been immediately omitted, and the highest words are the most often used.

Figure 30 - Word cloud visualisation



3.9 Model Variables

Features are the basic building blocks of datasets. There are three class labels we will predict for our tweets:

1. negative,
2. neutral or
3. positive

Figure 31 - Subjectivity and Polarity

	Tweets	Subjectivity	Polarity
0	hie solicitor sorry for the delay in respondin...	1.000000	-0.500000
1	kwazivai mavhima tinehurombo kunzwa izvi tinok...	0.000000	0.000000
2	sorry ishmaelchiwayu were it a wallet to bank ...	0.841667	0.200000
3	hie roygono our apologies for the experience k...	0.825000	0.500000
4	makadini lewistrap tinokumbirawo kuti mutarise...	0.000000	0.000000
...	...	...	...
195	hi tawandamaching laelliestrator our apologies...	0.900000	0.600000
196	mamukasei andrewthebe trixtar tinokutendai nek...	0.000000	0.000000
197	hi babaselma nhamo thank you for contacting us...	0.416667	0.116667
198	tinehurombo nekumanikidzika kwamasangana nako ...	0.000000	0.000000
199	hey kjiriyengwa we have responded pzp	0.000000	0.000000

200 rows × 3 columns

From subjectivity and polarity values above, while developing the model, the researcher was able to calculate the class labels for all the tweets

Figure 32 - Model variables

	id	myTweets	tweet_date	Subjectivity	Polarity	Analysis
0	1380450944861749255	hey katukablessing we apologise for the experi...	2021-04-09 09:22:07	0.9	0.6	Positive
1	1380449810696843265	zoeydaka hey nyaks we apologise for the experi...	2021-04-09 09:17:37	0.9	0.6	Positive
2	1380447806268268545	hie tinotendamutam we are sorry about that ple...	2021-04-09 09:09:39	1.0	-0.5	Negative
3	1380447594476937217	hie tinotendamutam we are sorry about that ple...	2021-04-09 09:08:48	1.0	-0.5	Negative
4	1380446238747803648	hi mbonkomo please check your dm for our respo...	2021-04-09 09:03:25	0.0	0.0	Neutral

3.10 Reliability and validity

To ensure the validity and reliability of the research instrument , detecting emotion from words yields better results if the subjects under study are not aware of their participation. This removes bias that comes with instruments such as questionaries and surveys. The researcher collected primary data in its raw form from twitter. The engagements between service provider agents and customers are unfiltered which validates the validity and reliability of the research. While more data sources such as incident management systems or call centre agent call logs can be incorporated , Twitter offers valid and reliable data that can be used in various other researches pertaining to sentiment analysis. Validity refers to how much the information gathered is relevant to the inquiry (Ghauri and Gronhaug, 2005). Validity is described as "measuring what is supposed to be measured" (Field, 2005). The degree to which a calculation of a phenomenon produces a stable and consistent outcome is referred to as reliability (Carmines and Zeller, 1979). The term "reliability" also refers to the capacity to replicate something. A scale or test, for example, is said to be accurate if repeated measurements taken under the same conditions yield the same result (Moser and Kalton, 1989). The researcher used the Test-retest approach of his methodology to ensure that the sample was accurate. The same instrument was applied repeatedly to the same twitter handle at different times to 4 different datasets.

3.11 Ethical considerations

Ethical considerations are one of the major considerations of any research study. Crosswell (2003) asserts that when conducting a study, the researcher has the mandate to observe the desires, needs, rights and values of participants. It therefore means that respondents have the ultimate right to making reasonable decisions in as far as responses are concerned (Graziano and Raulin, 2004) as well as ensuring that the respondents identify the findings of the study as their experiences (Streubert and Carpenter, 2011). All users' personal identifying information was strictly protected according to Twitter's user privacy terms and all user identity-related text content from user tweets was not presented in any tables or graphics. The information gathered was considered for academic use only.

3.12 Section summary

In this chapter, attention was given to the crucial elements that make up the research methodology applied in this study. The next chapter, Results and discussion, the task was to present the data that has been collected and analysed using the instruments discussed in this chapter

Chapter 4: Results and discussion

4.1 Introduction

This chapter presents findings from the analysis of the data collected from Twitter handles of 4 different service providers.

Table 9 - Service provider twitter handles

Service Provider	Twitter Handle
Econet Wireless	@econet_support
Telone Pvt Ltd	@TelOneZW
Zimbabwe Online (ZOL)	@ZOLconnect
Netone Pvt Ltd	@NetOneCellular

The presentation and analysis focused on the research objectives which were raised in chapter one in line with the research questions. After cleaning the scrapped user tweets using python and pandas ,a lexicon based sentiment analysis model was used to analyse the data to extract the polarity , subjectivity to determine the underlying emotion of customers. 200 user tweets per request were obtained from the above twitter handles . Data was analysed using Python. There are various sentiment analysis methods which were analysed by the researcher during literature review of which comparative studies of these methods have been carried out by various researchers. Results from these studies show that the results are somehow similar in accuracy with small differences observed when a combination of these methods are used. Henceforth the researcher focused more on the application of sentiment analysis as a new concept in Zimbabwe for evaluating customer satisfaction.

4.2 Model Development

The method of evaluating the author's mood or emotion, whether positive, negative, or neutral, is known as sentiment analysis.
The sentiment property of text blob returns two properties, polarity and subjectivity, by using the dataset below.

Figure 33 - Sample dataset showing subjectivity and polarity

	id	myTweets	tweet_date	Subjectivity	Polarity
0	1365589936087400448	Hi @MuleyaElijah, we have responded. ^MBM	2021-02-27 09:09:46	0.000000	0.000000
1	1365589888448479235	Hie @JamesMaseko7. Kindly note that we have re...	2021-02-27 09:09:35	0.900000	0.600000
2	1365589596344627201	Hie @JamesMaseko7, our apologies for the delay...	2021-02-27 09:08:25	0.000000	0.000000
3	1365587528900870144	Hello @Ajuda47631912, the bundles work like an...	2021-02-27 09:00:12	0.541667	0.083333
4	1365586710344708099	Makadini @TafaraPalesa. Tinokumbirawo mutarise...	2021-02-27 08:56:57	0.000000	0.000000

Subjective sentences generally refer to opinion, emotion or judgment whereas objective refers to information. Subjectivity is also a float which lies in the range of [0,1].

Polarity is a float in the form [-1,1], with 1 denoting a positive statement and -1 denoting a negative statement. In sentiment analysis, polarity refers to distinguishing positive, neutral, and negative sentiment orientation in written or spoken language.

Below is the function used to compute emotion from polarity.

Figure 34 - Computing emotion from polarity

Create a function to compute negative (-1), neutral (0) and positive (+1) analysis

```
def getAnalysis(score):
    if score < 0:
      return 'Negative'
    elif score == 0:
      return 'Neutral'
    else:
      return 'Positive'

df['Analysis'] = df['Polarity'].apply(getAnalysis)

# Show the dataframe
df.head()
```

Below are sample results after running the above function on the extracted dataset.

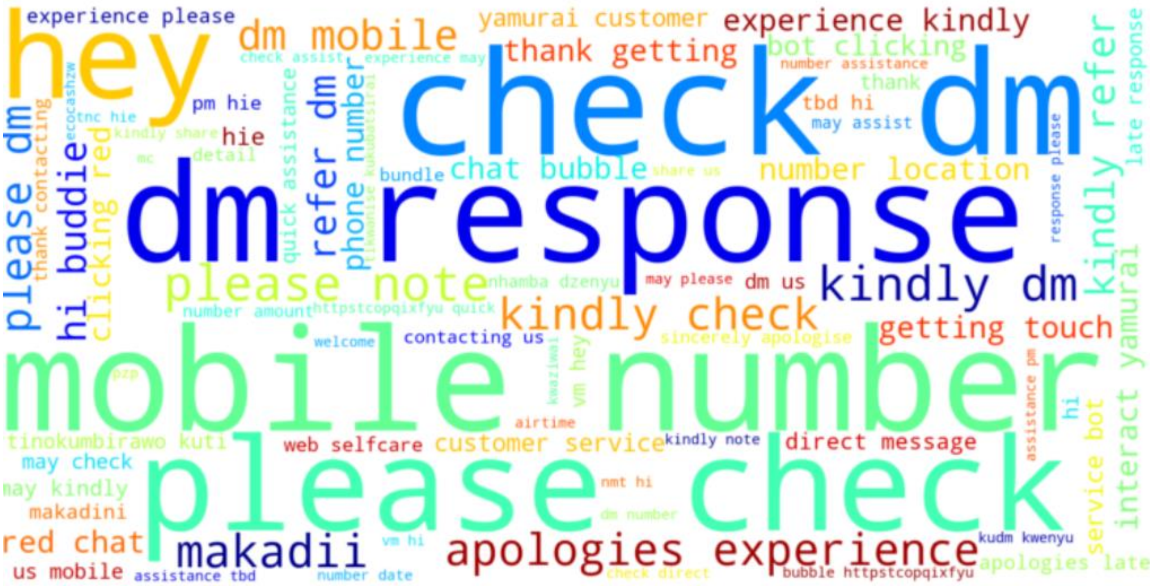
Figure 35 - Sentiment detection

	id	myTweets	tweet_date	Subjectivity	Polarity	Analysis
0	1365589936087400448	Hi @MuleyaElijah, we have responded. ^MBM	2021-02-27 09:09:46	0.000000	0.000000	Neutral
1	1365589888448479235	Hie @JamesMaseko7. Kindly note that we have re...	2021-02-27 09:09:35	0.900000	0.600000	Positive
2	1365589596344627201	Hie @JamesMaseko7, our apologies for the delay...	2021-02-27 09:08:25	0.000000	0.000000	Neutral
3	1365587528900870144	Hello @Ajuda47631912, the bundles work like an...	2021-02-27 09:00:12	0.541667	0.083333	Positive
4	1365586710344708099	Makadini @TafaraPalesa. Tinokumbirawo mutarise...	2021-02-27 08:56:57	0.000000	0.000000	Neutral

The analysis of subjectivity and polarity was done on all 4 service provider datasets collected on twitter. Below are model results for all service providers

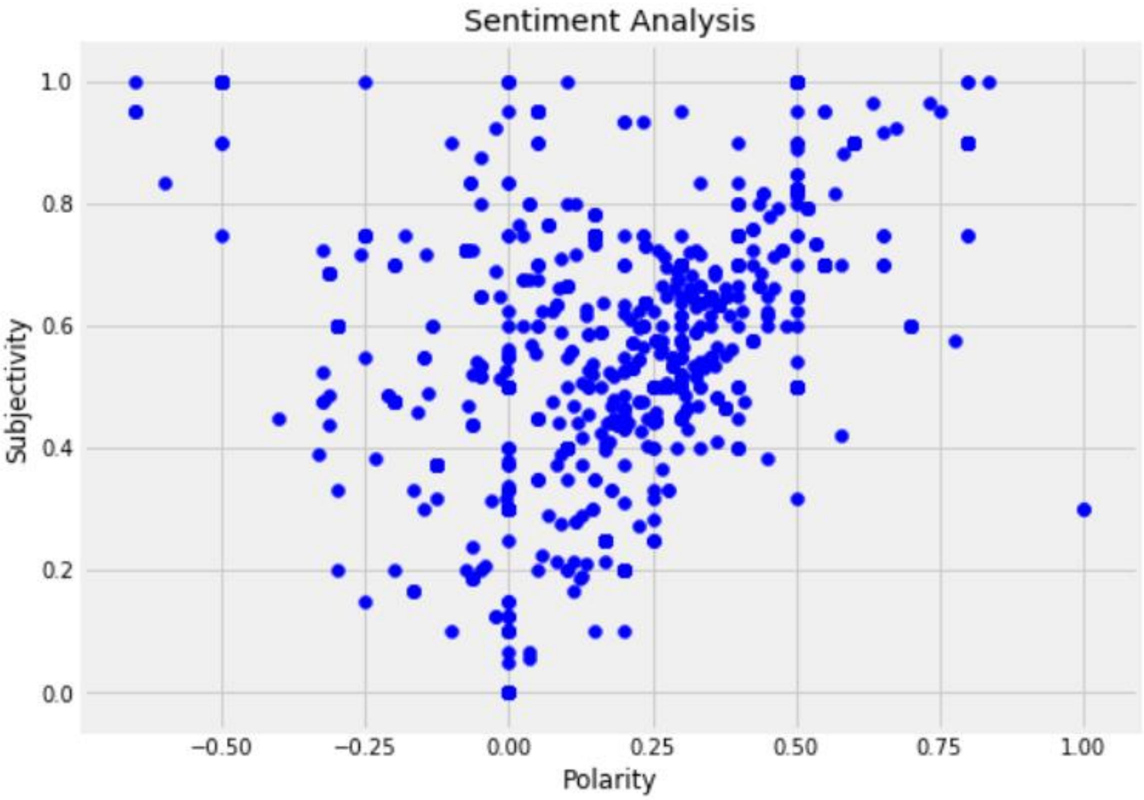
4.2.1 Model results for Econet Wireless Zimbabwe

Figure 36 - Econet word cloud visualisation



The word cloud above is showing us that the most used words used were DM ,response , Check ,please check and mobile number. Mobile and internet service providers can use this information to advise their customer service agents on the words to use which can greatly improve customer engagements.

Econet customer satisfaction Index



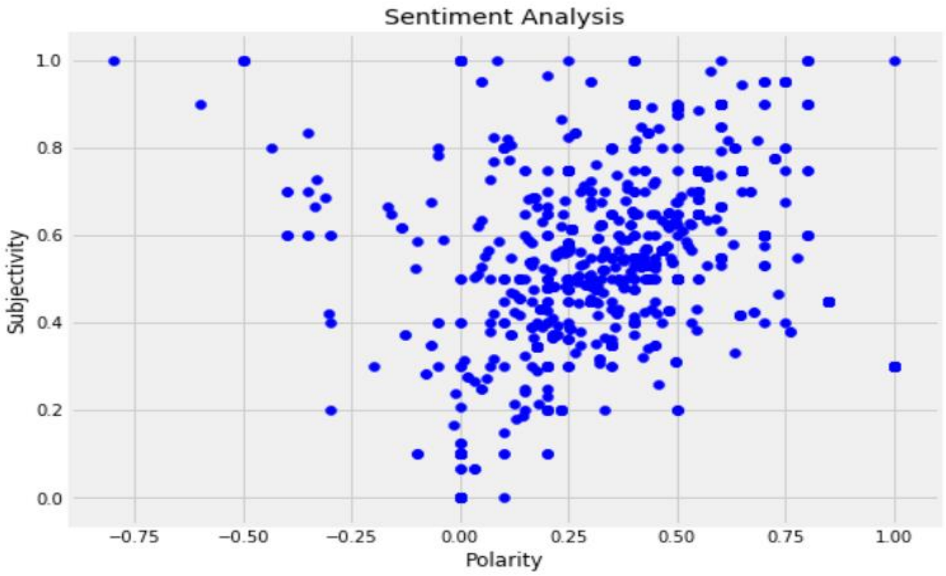
A simple look there are more dots on the positive side of polarity, meaning that the general engagement that econet does with its customers are generally more positive than negative. The researcher concluded that since most companies have embraced social media , customer services agents watch what they say.

4.2.2 Model results for Zimbabwe Online

Figure 37 - ZOL word cloud visualisation

The word cloud above is showing us that the most used words used were contacting , us ,good ,day , thank and please. Mobile and internet service providers can use this information to advise their customer service agents on the words to use which can greatly improve customer engagements.

Zimbabwe Online customer satisfaction Index -



A simple look there are more dots on the positive side of polarity, meaning that the general engagement that econet does with its customers are generally more positive than negative. The researcher concluded that since most companies have embraced social media , customer services agents watch what they say.

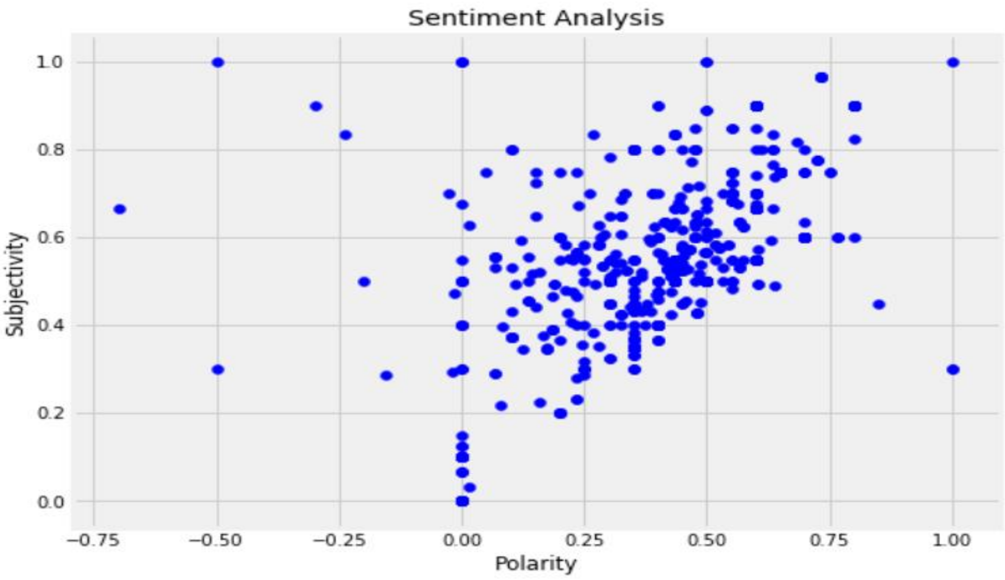
4.2.3 Model results for Telone Pvt Ltd

Figure 38 - Telone word cloud visualisation

The word cloud above is showing us that the most used words used were kindly ,response , maskup , evening. Mobile and internet service providers can use this information to advise their customer service agents on the words to use which can greatly improve customer engagements.

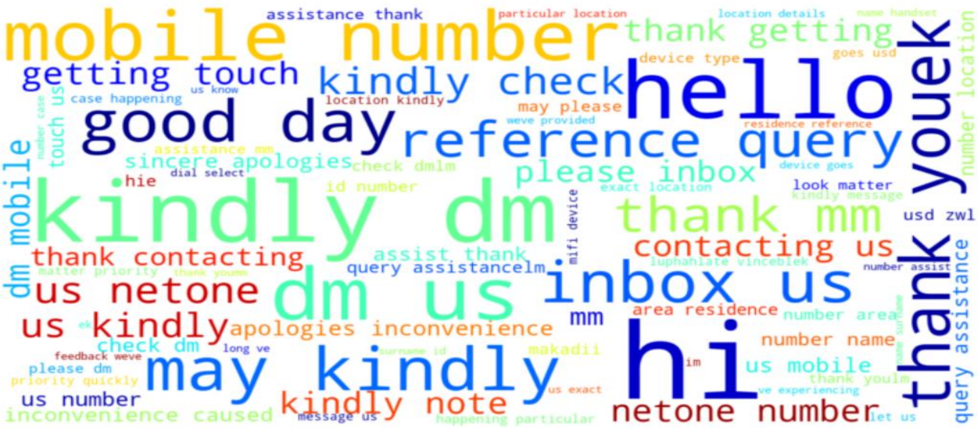
Telone customer satisfaction Index –

A simple look there are more dots on the positive side of polarity, meaning that the general engagement that econet does with its customers are generally more positive than negative. The researcher concluded that since most companies have embraced social media , customer services agents watch what they say.



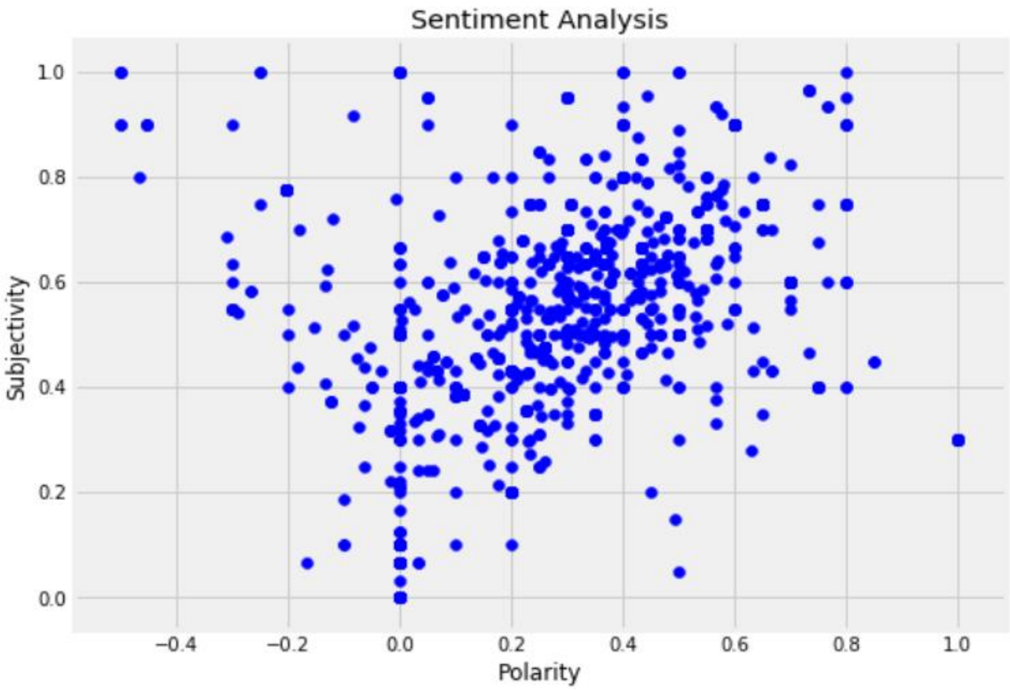
4.2.4 Model results for Netone Pvt Ltd

Figure 39 - Netone word cloud visualisation



The word cloud above is showing us that the most used words used were hi ,mobile , number ,kindly and inbox. Mobile and internet service providers can use this information to advise their customer service agents on the words to use which can greatly improve customer engagements.

Netone customer satisfaction Index -



A simple look there are more dots on the positive side of polarity, meaning that the general engagement that econet does with its customers are generally more positive than negative. The researcher concluded that since most companies have embraced social media , customer services agents watch what they say.

4.3 General analysis of the results for all 4 providers

All 4 providers are heavy on the neutral line and reason being customer services representatives are expected to answer as professional as they can.

4.3.1 Neutral Tweets

```
'5) hi mbonkomo please check your dm for our response hm'

'7) nc prikllar chunkieyrussell hatiperiwacho ricardothezimbo theavengerzw axecuresneaker ayiiejuicy kdotjnrwurlwid
e abdesigns rega tifambire nyaya dzacho ha makaoma henyu mc'

'8) chunkieyrussell prikllar hatiperiwacho ricardothezimbo theavengerzw axecuresneaker ayiiejuicy kdotjnrwurlwide a
bdesigns hi russellmuface hatisi kumuziva uyu kkk mc'

'9) makadini gonorutendo takusiyirai mhinduro kudm kwenyu hm'

'11) harry thats alright we will be in touch with you soon ken'

'12) hi titobambo please share with us the buying number amount transaction date billing channel ecocashmain airtim
e and your mobile number  hm'

'14) hi gladysm please refer to your dm hm'

'15) please check your dm for our response gladysm hm'

'16) hi dexterschase please share with us the buying and receiving number amount and transaction date for further a
ssistance hm'
```

Neutral words are defined as all those words that are neither positive nor negative. Words with scores near zero could be thought of as neutral words. Looking at the extracted sample of neutral tweets , we can tell that the meaning of the sentence when put into context do not add value to a customer. Service providers can analyse these sentences and find ways to turn them into positive sentiment. e.g., looking at this sentence extracted from the except above

'15) please check your dm for our response gladysm hm' – this can be rephrased by customer service gents to

'15) We have resolved your issue. please check your dm for our response gladysm hm'

4.3.2 Negative tweets

Figure 40 - Negative Tweets

```
1) hie tinotendamutam we are sorry about that please check your dm for our response kk

2) hie tinotendamutam we are sorry about that please check your dm for our response kk

3) we sincerely apologize for the terrible experience andrewbakuza please check your dm for our response cfm

4) our apologies shammahbupe for the unpleasant experience the issue has been resolved nza

5) hi joshychibuda gracie thank you for contacting us please note the private wifi bundle has a window period of  d
ays and if you fail to utilize it within the stipulated time you are unable to recover it tbd

6) hie ngtemba we are here to assist may you please check our response to your other tweet tfm

7) hi zandile were sorry about that may we have the date of purchase for further assistance nmt

8) hi cdetatonga were sorry about that please dm your phone number and the date of purchase for further assistance
nmt
```

No matter how successful the company is, there will be moments when you fall short of your customers' standards. Any business offering a service or selling a product receives bad reviews from time to time. Negative sentiment may assist service companies in paying attention to the problems that their clients are having with their services. Negative sentiment can be highly damaging to the brand's image if ignored or suppressed. Customers who had especially bad experiences will be identified, as well as why and how they can be helped. They can also tell the difference between regular feedback and feedback that influences satisfaction ratings.

4.4 Feasibility or applicability of model

The research was carried out within the context of the Zimbabwean setting where twitter engagements between internet service providers and customers would mix english and native languages Shona and Ndebele. The resultant model however can be applied to any english text , detect and classify sentiment.

4.5 Future work and Limitations

For future work ,it will be ideal to incorporate a linguistic model that can either translate the native language text to english or interpret the native languages during sentiment analysis.

4.6 Implications of the model

The advent of social media has spawn new sources of data where organisations are now experiencing data overload . While this does not translate to better or deeper insights, organisations are now equipped to collect huge amounts of customer feedback. It is still difficult to analyse it without any sort of error or bias. This is where sentiment analysis comes into play. Data from various sources , mostly social media can be streamed in real time , stored in memory or on cloud storage for analysis later. Sentiment analysis provides sentiment scores that are comparable to, and potentially more refined than star ratings or NPS scores. Customer sentiment can be tracked over time to gain insight into why net performance scores (NPS) or sentiment against specific facets of the market have improved. Though sentiment is ubiquitous as a term, quantifying sentiment or concepts like emotionality, negativity, polarity, subjectivity, sound, or valence is difficult. In the social sciences, there is a lack of agreed-upon conceptualization and operationalization, whereas sentiment is commonly articulated with vague and inventive vocabulary (Liu, 2012; Pang & Lee, 2008; Wiebe et al., 2004).

4.7 Chapter summary

This Chapter looked at the results of various analysis using word cloud and bar charts. Data was presented, analysed, and interpreted in line with the objectives of the study which sought to measure customer satisfaction , extract the most prevalent words and develop a twitter sentiment predictive model by analysing customer sentiment from tweets.

Chapter 5: Summary, Conclusions and Implications

5.1 Introduction

This chapter concludes the research study. Answers to research questions raised in the first chapter will be outlined and conclusions will be clearly stated. For future studies, recommendations are going to be provided. The recommendations and suggestions are based on the key findings obtained in the previous chapter and the objectives of the study.

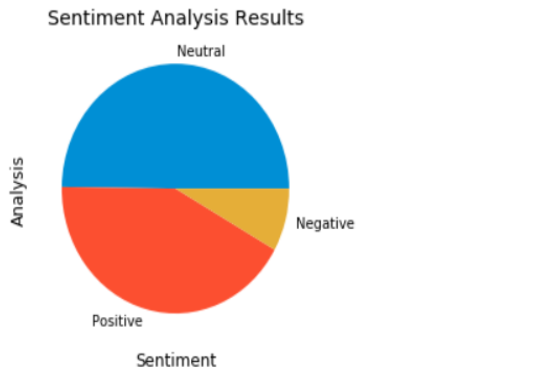
5.2 Summary of research findings

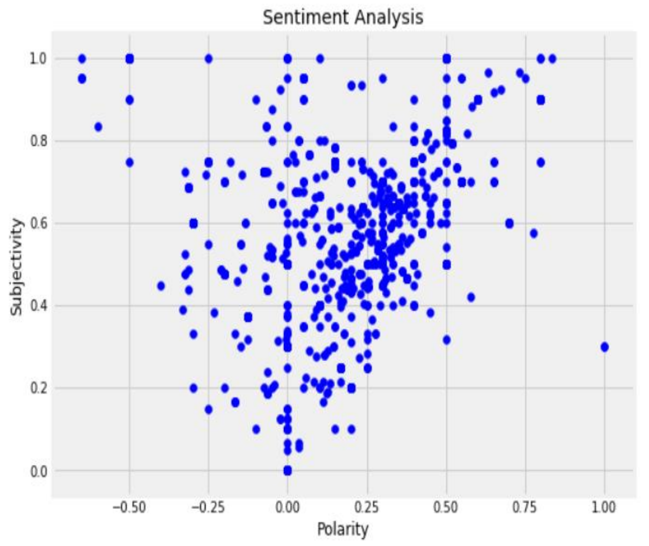
Based on the objectives of the study the research was able to meet the objectives of the study and answered all research questions.

Research Objective	**Research Question**
To measure customer satisfaction by analysing customer sentiment from tweets	How do we measure customer satisfaction using sentiment analysis?

Using the pie chart and scatter graph below , the researcher managed to satisfy the above research objective.

```python
# Plotting and visualizing the counts
plt.title('Sentiment Analysis Results')
plt.xlabel('Sentiment')
plt.ylabel('Counts')
df['Analysis'].value_counts().plot(kind = 'pie')
plt.show()
```





Research Objective | **Research Question**

Research Objective	Research Question
To extract the most prevalent words that contribute to a negative or positive sentiment.	What words can be used to improve customer engagement?

Using the below word cloud visualisation , analysis was run against responses that were classified as positive or negative. Specific words were identified which could be used to educate customer service agents on how to respond.

```python
word_cloud(df['myTweets'])
```

Research Objective	**Research Question**
To identify the most suitable words that can be used to improve customer engagement.	What words can be used to improve customer engagement?

Carrying over the results of the word cloud visualisations , the researcher was able to determine how specific responses could be changed to better communicate the position of the business. Every business is always asking and looking for an answer to “How can we improve overall customer engagements and Influence customer behaviour?” Using the sample dataset below.

Randomly sample 10% of your dataframe

```
datasetC = df.sample(frac=0.1)
```

```
datasetC
```

	id	myTweets	tweet_date	Subjectivity	Polarity	Analysis
2012	1362681659330605059	hi docmoodsie elsamplero kindly refer to your ...	2021-02-19 08:33:19	0.900000	0.600000	Positive
2239	1362207539115552769	hie babangwenya kindly check our dm response gfk	2021-02-18 01:09:20	0.900000	0.600000	Positive
463	1365216000573079552	hey tendainherera we sincerely apologise for t...	2021-02-26 08:23:53	0.700000	0.000000	Neutral
2128	1362393042490122242	wakuraa hey buddie please check your dm for ou...	2021-02-18 13:26:28	0.000000	0.000000	Neutral
1455	1363833277044428804	hie tinodzikiti kindly confirm if the provided...	2021-02-22 12:49:26	0.700000	0.300000	Positive
2044	1362663874487586816	makoreaddmore hi buddie merchant reversals are...	2021-02-19 07:22:39	0.888889	0.500000	Positive

Randomly choosing 10% of the dataset and running the sentiment analysis model 4 times gave us 4 different results. Which proved that certain words can be identified and used to improve customer engagements. Moreover

Research Objective	**Research Question**
To develop a twitter sentiment predictive model	How do we develop a sentiment analysis model ?

Running the following code snippet against polarity of the dataset, the researcher developed a lexicon based model which can be applied on any dataset.

```
def getAnalysis(score):
    if score < 0:
      return 'Negative'
    elif score == 0:
      return 'Neutral'
    else:
      return 'Positive'

df['Analysis'] = df['Polarity'].apply(getAnalysis)

# Show the dataframe
df.head()
```

	id	myTweets	tweet_date	Subjectivity	Polarity	Analysis
0	1365672925534294016	hi tiniezoey kindly refer to your dm for our r...	2021-02-27 14:39:33	0.90000	0.60000	Positive
1	1365671143294787595	makadini giftvocure chagumukaa demarcusmhalo t...	2021-02-27 14:32:28	0.00000	0.00000	Neutral
2	1365670345773092866	makadini chagumukaa takupindurai kudm kwenyu mbm	2021-02-27 14:29:18	0.00000	0.00000	Neutral
3	1365667945372860426	hie maiteeteebi we are glad that the issue has...	2021-02-27 14:19:45	0.81875	0.44375	Positive
4	1365667876544393218	hi ultimatekagz demarcusmhalo kindly refer to ...	2021-02-27 14:19:29	0.90000	0.60000	Positive

5.3 Practical implications

This study may lead businesses in Zimbabwe to incorporate sentiment analysis techniques when evaluating customer satisfaction as it provides more insights as compared to old traditional methods. Tapping into social media has provided significant benefits for many organisations. However, most of them are only monitoring social media without collecting that data for further analysis. If these insights are acted upon, organisations will improve their customer experience and in turn their revenues significantly since a satisfied customer is a loyal customer. Advantages of using customer satisfaction metrics obtained from sentiment analysis

5.3.1 Loyal customers

Customers who are happy stay loyal and spend more money. They always come back because they like a product or service and are happy with the results and customer service they received. These metrics will help you identify your most loyal and happy consumers. You should reach out to unhappy people and learn about their motivations for dissatisfaction and take steps to make them happier.

5.3.2 Promoters

You will find out who the promoters, opponents, and passives are by using NPS surveys. You will use your promoters to amplify your brand and optimistic aspects. They are the best brand ambassadors, and you can use them to your advantage. You will nudge passives into being promoters by taking steps. Discounts, coupons, early product releases, and other perks may be used to do this. It would be beneficial if you focused more on opponents, getting to the root of their frustration and negative experience. This would aim to boost the overall customer experience and loyalty.

5.3.3 Brand reputation

Customers who have had positive experiences may tell their friends and family about it, while those who have had negative experiences will tell their friends and family about it. This not only harms the brand's integrity, but it also hurts the bottom line. CSAT surveys will help ease some of these issues by allowing you to monitor and make amends with these consumers. Users will quickly upload information about their less than favourable interactions or poor feedback when there are so many social media platforms at their fingertips. You will prevent this by acting quickly and taking the steps required to increase their experience and happiness.

5.3.4 Usability and experience

You can gather quantitative and qualitative data from your customers by conducting Customer Effort Score (CES) surveys. You should use these experiments to make sure you're making their lives better and keep that in mind when you carry out new features or services. All these metrics have their own set of advantages that can be used to improve customer loyalty.

5.4 Conclusions

The popularity of sentiment analysis is soaring. However, even if the field is developing, it is still fresh, and the researcher may face difficulties. The essence of grouping is one potential stumbling block. The number of groups and subgroups that can be derived is generally limited, with most classification techniques producing just two to three groups at most. Text-based data is often context and domain dependent, accurate only in specific locations and at specific times. While there might be some localization, mistranslation may jeopardize the authenticity of the translated document. Furthermore, since they often represent shorter versions of sentences, internet posts can be challenging to analyze. Other critiques of sentiment analysis are directed at the methods used.Machine learning, for example, assigns classifications based on the score created from a data corpus. The development of this can be very costly and time-consuming. Furthermore, overall classification precision is dependent on classification results, which may or may not be transferable to other domains. The accuracy of automatic emotion classification was called into question in a recent article published by the Academy of Marketing Science (Davis and O'Flaherty, 2012). Automatic coding firms were more likely to misclassify emotions when sentences were lengthy, did not contain keywords or subject phrases, and had reversed definitions due to the negation effect, according to the authors. Sentiment classification, on the other hand, proved to be highly reliable (80% and higher) when sentences were simple, and the sentiment polarity was obvious. Finally, ethical research concerns can arise. In this study, the right to voluntary involvement, secrecy, and confidentiality can be called into question, provided that the data is extracted and analysed without the author's permission. In conclusion, emotion analysis is a comparatively recent concept in the field of science. Nonetheless, contributing to the real-time translation of large amounts of textual data into usable information can be extremely beneficial. The cost, time, and processing advantages are sufficient to justify this academic focus. When the number of people purchasing, consuming, and conversing online has increased, so has the number of marketers tasked with sifting through online textual content. ASSANA (Accelerating Social Sciences for the Modern Age) is a new initiative aimed at discovering, improving, and disseminating new social science methodologies (Cougburn, Hansen and Wozniac, 2012). This project acknowledges the difficulties that social science researchers encounter when analysing textual evidence. It encourages social scientists to use digital evidence in their study. Sentiment research may be beneficial. Sentiment analysis allows market researchers to gather deep, rich qualitative data from a vast number of participants in an unobstructed real-world context without external interferences. It also offers a comprehensive method for collecting and interpreting vast amounts of text in real time. It eliminates subjectivity and personal prejudice. In this demanding and modern environment, sentiment analysis offers a systematic and thorough methodology for interpreting evidence. It has the potential to close the divide between qualitative and quantitative research debates and offer a deeper, more integrated insight into online market research when properly integrated with current research design methods.

5.5 Recommendations - Future research implications

5.5.0 Add domain specific features in the Corpus

The study's results highlighted a void in the current literature on the use of emotion analysis in Zimbabwean studies. The findings may lead to further research into sentiment analysis linguistic models, especially those that include native languages like Shona and Ndebele. There were a couple of neutral emotions in the data scraped from Twitter for both services, which affected the accuracy of the satisfaction ratings. Adding domain-specific functionality to the Corpus is one way to enhance this.

5.5.1 Use an exhaustive stopwords List

The most often used words in a corpus are known as stopwords. “A, the, of, on, etc.” are the most often used stopwords. These terms are used to describe a sentence's structure. However, they are useless in describing the meaning. Treating these type of words as feature words would result in poor performance in text classification. To improve results, these words may be removed from the corpus entirely. Aside from language stopwords, there are a few other supporting words that are less relevant than the phrases themselves. There are some of them:

- Language Stopwords – a, of, on, the … etc
- Location Stopwords – Country names, Cities names etc
- Time Stopwords – Name of the months and days (January, February, Monday, Tuesday, today, tomorrow …) etc
- Numerals Stopwords – Words describing numerical terms (hundred, thousand, … etc)

5.5.2 Noise free corpus

In most data science questions, using a classification algorithm on a cleaned corpus rather than a noisy corpus is preferred. Unimportant text entities, such as punctuation marks, numerical values, links and urls, are referred to as noisy corpus. Since the size of the sampling space of available features sets shrinks as these entities are removed from the code, consistency improves. However, it's important to remember that these entities can only be excluded if the classification issue doesn't use them.

5.5.3 Use Complex Features: n-grams and part of speech tags

In certain cases, considering features as a group of words has more meaning than considering single words as features. N-grams are made up of N terms put together. Bigrams are well-known as the most descriptive N-Gram combinations. The consistency of the text classification model will be improved by using bigrams in the feature set. Similarly, combining Part of Speech tags with words/n-grams would have an additional collection of features. Increase the number of classifications as well. e.g., it is preferable to train the model such that the word "book" means "book of pages" when used as a NOUN and "book a ticket or something else" when used as a VERB.

Figure 41 - Word features

Old features	Modified features
Love	Love buddie
Bundle	Bundle depleted
Internet	Slow internet
Speed	Slow speed

5.4 Chapter summary

This chapter presented the conclusions drawn from this study about evaluating customer satisfaction using sentiment analysis within service providers. Recommendations were also made on how certain considerations can be made to improve the accuracy of the sentiment analysis model.

6 References

- Al-Otaibi, S., Alnassar, A., Alshahrani, A., Al-Mubarak, A., Albugami, S., Almutiri, N., & Albugami, A. (2018). Customer Satisfaction Measurement using Sentiment Analysis*. International Journal of Advanced Computer Science and Applications*: IJACSA, 9(2). doi:10.14569/ijacsa.2018.090216
- Anastasia, S., & Budi, I. (2016). Twitter sentiment analysis of online transportation service providers*. 2016 International Conference on Advanced Computer Science and Information Systems*. (ICACSIS), 359–365. IEEE.
- Berry, L. L. and Carbone, L. P. (2007). *Build loyalty through experience management*. Quality Progress40(9): 26-32.
- Biza, P. (2019). *Factors that affect customer satisfaction in telecommunications industry. A case of Telone Bindura*. BUSE.
- Cajetan, C. (2018). Digital banking, customer experience and bank financial performance: UK customers' perceptions. *International Journal of Bank Marketing. Vol. 36 No. 2*, pp. 230-255.
- Carbone, L. P. and Haeckel, S. H. (1994). *Engineering Customer Experience*. Marketing Management 3(3): 8-19
- D'Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, 125(3), 26–33.
- De Mauro, A., Greco, M. and Grimaldi, M. (2016). *A formal definition of Big Data based on its essential features*. Library Review, Vol. 65 No. 3, pp. 122-135. https://doi.org/10.1108/LR-06-2015-0061.
- Gate, T. (2017). *Investigation on the impact of customer relationship management on customer retentio*n. BUSE.
- Gitto, S., & Mancuso, P. (2017). *Improving airport services using sentiment analysis of the websites*. Tourism Management Perspectives, 22, 132–136.
- GUY, R. 2012. The use of social media for academic practice: A review of literature. *Kentucky Journal of Higher Education Policy and Practice*, 1, 7.
- Gwambuka, W. (2017). An assessment of the contribution of internet banking to customer satisfaction in the banking sector in Zimbabwe Retrieved February 20, 2021, from Buse.ac.zw:8080 website: http://liboasis.buse.ac.zw:8080/xmlui/handle/123456789/5872.
- Holmlund, M., Van Vaerenbergh, Y., Ciuchita, R., Ravald, A., Sarantopoulos, P., Ordenes, F. V., & Zaki, M. (2020). Customer experience management in the age of

big data analytics: A strategic framework. *Journal of Business Research*, 116, 356–365.

- Homburg, C., Jozié, D., & Kuehnl, C. (2015). Customer experience management: toward implementing an evolving marketing concept. *Journal of the Academy of Marketing Science*, 45(3), 377–401.
- http://www.potraz.gov.zw/wp-content/uploads/2017/03/Annual_Sector_Perfomance_Report_2017.pdf
- https://sproutsocial.com/insights/listening-vs-monitoring/
- Hwambo, L., Shamhuyenhanzva, R. M., & Sandada, M. (2017). An assessment of churn determinants in Zimbabwe's mobile telecommunication services*. International Journal of EBusiness and EGovernment Studies*, 9(2), 106–120.
- J. Carvalho, A. Prado and A. Plastino,. (2014*). A Statistical and Evolutionary Approach to Sentiment Analysis*. IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Warsaw, 2014, pp. 110-117, doi: 10.1109/WI-IAT.2014.87.
- Jena, R. (2017). An evaluation of the contribution of Information and Communications Technology (ICT) to customer satisfaction in the banking sector in Zimbabwe: A case study of MBCA Bank (2010 - 2013). Retrieved from http://41.175.146.201/handle/10646/3595.
- Kaplan, A.M., Heinlein, M. (2010). *Users of the world, unite! The challenges and opportunities of social media*. Bus Horiz 53(1):59–68.
- Kaplan, A. M. & Heinlein, M. 2010. *Users of the world, unite! The challenges and opportunities of Social Media.* Business horizons, 53, 59-68.
- Kapoulas, A. & Mitic, M. 2012. Understanding challenges of qualitative research: rhetorical issues and reality traps. *Qualitative Market Research: An International Journal*, 15, 354-368.
- Karim, R.& Chowdhury, T. (2014). Customer satisfaction on service quality in private commercial banking sector in Bangladesh*. British Journal of Marketing Studies*. 2. 1-11.
- Keiningham, T., Joan B., Sabine (ne´e Moeller) B., Helen L., Alexander B., Julija, D., Linda, N., OuYi-Chun., and Zaki, M. (2017). The Interplay of Customer24Journal of Service Research 22(1) Experience and Commitment, *Journal of Services Marketing*,31(2), 148-160.
- Komariah, K. S., Machbub, C., Prihatmanto, A. S., & Sin, B.-K. (2016). A Study on Efficient Market Hypothesis to Predict Exchange Rate Trends Using Sentiment

Analysis of Twitter Data. *Journal of the Korean Multimedia Society*, 19(7), 1107–1115. https://doi.org/10.9717/KMMS.2016.19.7.1107

- Laroche, M., Habibi, M. R. & Richard, M.O. (2013). To be or not to be in social media: How brand loyalty is affected by social media? *International Journal of Information Management*, 33, 76-82.
- Lemon, Katherine N. and Peter C. Verhoef. (2016). Understanding customer experience throughout the customer journey. *Journal of marketing*. 80 (6), 69-96
- Leung, D., Law, R., Van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing, 30*(1-2), 3-22.
- Levinson, J. C. & Perry, D. E. (2011). *Guerrilla marketing for job hunters 3.0: How to stand out from the crowd and tap into the hidden job market using social media and 999 other tactics today*. John Wiley & Sons.
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Williston, VT: Morgan & Claypool Publishers.
- Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*: Cambridge University Press.
- Macdonald, Emma K., Michael Kleinaltenkamp, and Hugh N. Wilson. (2016). How business customers judge solutions: solution quality and value in use. *Journal of Marketing, 80 (3), 96-120.*
- Makanyeza, C., & Chikazhe, L. (2017). Mediators of the relationship between service quality and customer loyalty: Evidence from the banking sector in Zimbabwe. *International Journal of Bank Marketing*, 35(3), 540–556.
- Makudza, F. (2020). Augmenting customer loyalty through customer experience management in the banking industry. *Journal of Asian Business and Economic Studies*
- Market Research. (2020, November 14). In *Wikipedia*. https://en.wikipedia.org/wiki/Market_research
- Marumbwa, J. and Mutsikiwa, M. (2013) 'An Analysis of the Factors Influencing Consumers' Adoption of mobile money transfer services (mmts) in Masvingo urban, Zimbabwe', *British journal of economics, management & trade, Vol. 3(4), pp. 498-512.*
- Mayes, L. (2011). Effectively incorporating social media: A Case Study on Coca-Cola. Expedition, 206, 27.
- Mazikana, A. T. (2020). The impact of quality service and customer satisfaction on consumer purchasing decisions in Zimbabwe telecommunications industry. SSRN Electronic Journal. doi:10.2139/ssrn.3715253.

- McCall, T. (2015). Gartner predicts a customer experience battlefield. Retrieved from https://www.gartner.com/smarterwithgartner/customer-experience-battlefield/. Accessed 27 December 2020.
- McColl, K., Janet R., Stephen L. Vargo., Tracey S. D., Jillian C. S., and Yasmin van Kasteren. (2012). Health care customer value correlation practice styles. *Journal of Service Research*, 15 (4), 370-389.
- Medhat, W., Hassan, A. & Korashy, H. (2014). *Sentiment analysis algorithms and applications*: A survey, pp. 1093-1113, Vol. 5.
- Patino, A., Pitta, D.A. and Quinones, R. (2012). Social media's emerging importance in market research. *Journal of Consumer Marketing, Vol. 29 No. 3, pp. 233-237. https://doi.org/10.1108/07363761211221800*
- Perumal, S. (n.d.). Literature review on sentiment analysis. Retrieved November 29, 2020, from Ijstr.org website: http://www.ijstr.org/final-print/apr2020/Literature-Review-On-Sentiment-Analysis.pdf
- Postal and Telecommunications Regulatory Authority of Zimbabwe(POTRAZ). (2020). Abridged postal & telecommunications sector performance report. Retrieved from https://t3n9sm.c2.acecdn.net/wp-content/uploads/2020/12/Abridged-Sector-Performance-report-3rd-Q-2020.pdf
- POTRAZ, (2019), *Abridged Postal and Telecommunications Sector Performance Report Second Quarter 2019*
- Rambocas, Meena. (2013). *Marketing research: The role of sentiment analysis*. FEP WORKING PAPER SERIES.
- Ruddell, R. & Jones, N. 2013. *Social media and policing: matching the message to the audience*. Safer Communities, 12, 64-70.
- Shahin, A. and Janatyan, N. (2011) Estimation of Customer Dissatisfaction Based on Service Quality Gap by Correlation and Regression Analysis in Travel Agency. *International Journal of Business Management*, 6, 99-108.
- Stelzner, M. (2013). 2013 Social media marketing industry report: how marketers are using social media to grow their businesses, social media marketing examiner.
- Tsvetovat, M., Kazil, J., & Kouznetsov, A. (2012). Implicit sentiment mining-New approach overcomes inherent problems while exploring media bias and electoral sentiment. *OR/MS Today*, 39(6), 20.
- Twitter.Inc. Twitter privacy policy. 2017. https://twitter.com/en/privacy.
Google Scholar
- Tyagi, Priyanka and Chakraborty, Sudeshna and Tripathi, R.C. and Choudhury, Tanupriya, Literature Review of Sentiment Analysis Techniques for Microblogging

Site (March 15, 2019). International Conference on Advances in Engineering Science Management & Technology (ICAESMT) - 2019, Uttaranchal University, Dehradun, India, Available at SSRN: https://ssrn.com/abstract=3403968

- Verhoef, P. C., Kooge, E. & Walk, N., (2016), Creating Value with Big Data Analytics: Making Smarter Marketing Decisions. New York, NY: Routledge.
- Verhoef, P.C., Lemon K.N., Parasuraman, A.R., Michael T., & Leonard A. Schlesinger (2009). Customer Experience Creation: Determinants, Dynamics and Management Strategies. *Journal of Retailing*, 85 (1), 31-41.
- Villarroel O., Francisco, B.T., Jamie B., Thorst G. and Zaki M. (2014). Analysing customer experience feedback using text mining: a linguistics-based approach. *Journal of Service Research*, 17 (3), 278-295.
- Viriri, P., & Phiri, M. (2017). Determinants of Customer Satisfaction in Zimbabwe Telecommunication Industry. *Journal of Communication*. 8. 101-104. 10.1080/0976691X.2017.1317495.
- Vivek, S. D., Beatty, S. E., & Morgan, R. M. (2012). Customer engagement: Exploring customer relationships beyond purchase. *The Journal of Marketing Theory and Practice*, 20(2), 122–146.
- Voss, C. and Zomerdijk, L. (2007). Innovation in Experiential Services -An Empirical View. Innovation in services. London, DTI: 97-134.
- Zeithaml, V.A., Parasuraman, A. and Berry, M. (1990), Delivering Service Quality, The Free Press, New York.
- Zhou, L. & WANG, T. (2014). Social media: A new vehicle for city marketing in China. Cities, 37, 27-32.

Appendix 2: Approval Form and Declaration

Approval Form

The undersigned certify that they have read and recommended to the Graduate Business School, Chinhoyi University of Technology, for acceptance a dissertation entitled, **Evaluating customer satisfaction using sentiment analysis: A case study for mobile and fixed internet service providers in Zimbabwe**, submitted by Martin Chamambo, in partial fulfilment of the requirements for the Master of Science Degree in Data Analytics.

Name of Supervisor

Mr H Mazhokota

Signature

… …

Date :25/04/21

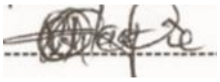
Declaration

I, Martin Chamambo, declare that this MSc study is my own effort and is a true reflection of research executed by me. This research in full or part thereof has not been submitted for examination for any degree at any other university/institution.

No part of this dissertation may be reproduced, stored in any retrieval system, or transmitted in any form, or by any means (e.g., electronic, mechanical, photocopying, recording or otherwise) without the prior express permission of me the author, or Chinhoyi University of Technology on my behalf.

I, Martin Chamambo, grant Chinhoyi University of Technology permission to reproduce this dissertation in whole or in part, in any manner or format, which Chinhoyi University of Technology may deem fit.

Name of Student : ……...Martin Chamambo

Signature : ……

Date : ……25/04/2021