Copenhagen Business School

Department of IT Management

Effect of Text mining techniques on Jabra Quality Management

Case of Consumer Generated Content

KAN-CINTO1005U-Master Dissertation

Master of Science (Msc.) in Business Administration and Information Systems

2017

Supervisor

Professor Ravi Kiran Vatrapu

Authors

Ngoc Ha Le

Palwasha Mahmood

Number of pages: 111

Number of characters: 174,488

Date of submission: May 15th, 2017

Acknowledgements

The completion of this dissertation could have not been accomplished without the warm support of many people of whom we are deeply grateful. First and foremost, we would like to express our gratitude to Prof. Ravi Vatrapu, our supervisor for this thesis, for his constant guidance and instruction throughout the research progress.

We are extensively indebted to the Copenhagen Business School, without whom none of our academic achievements would have been possible.

We profoundly appreciate the fact that Jabra trusted our abilities to run the experiment, by providing us with the data.

Last but not least, we are deeply thankful to our friends and family for their unconditional support, encouragement and care which all positively contributed to the accomplishment of the thesis.

Abstract

The primary objective of this paper is to demonstrate the ability of data analytics in capturing and translating consumer opinions into sentiment identification with the application of automation techniques. By applying various data analytic techniques, we propose an effective model to retrieve consumers' reviews from different platforms and transform them into useful piece of information to support the decision making process in organizations.

This study focuses on the aspects of social content generated by consumers. Web 2.0, and social media platforms have changed the behavior of consumers on the Internet. Consumers are now utilizing these social platforms to express their own experience about products and services. These opinions carry valuable information reflecting on consumer's response to the products. Consequently, there is a need for business to have an appropriate method to capture information from the source of data.

To address this concern, this paper aims to explore opportunities for automation in the consumer opinion mining and product quality assessment. The design framework is used in conjunction with literature review, theoretical analysis and creative method in order to answer the research question *"What is the effect of text mining techniques on quality management of Jabra, using consumers' product reviews?"*

We use the consumer reviews and opinions about Jabra from Amazon and Bestbuy for our data to do the analysis of product features based on the reviews. The main concepts we have used for our research include social data analytics, sentiment analysis and natural language processing. The dataset consists of consumer reviews extracted from the two dominant e-commerce websites, Amazon and Bestbuy. The main tools used were Python and its supported library (Textblob, NLTK, sklearn and statsmodel), SQLite and QlikSense

The result confirms the fact that social data consists of valuable insight from consumers and also for product quality control manager to improve product quality. Therefore, organizations need to focus on developing models based on these reviews to take advantage of the big social data.

Keywords: text mining, social data analysis, sentiment analysis, OLS regression analysis, product quality assessment, business intelligence, perceived quality

Table of Contents

List of figures

List of tables

About

Intended audience

The intended audience of this dissertation are individuals with a background in Information Technology and comparable to that. The findings of the research are valuable to both regular and online businesses. Knowledge of text mining methods, as that achieved by following an introductory course in information retrieval is expected. Training in Natural Language Processing (NLP) is preferable, but not a prerequisite. Last but not least, basic knowledge of statistical analysis helps in finding knowledgeable insights and patterns in the results.

Conventions used

The expressions

- "opinion mining" and "sentiment analysis"
- "customer reviews" and "consumer generated content"
- "consumer" and "customer"
- "key feature" and "key performance indicator"

will be used interchangeably to denote the purpose of determining the attitude of a speaker towards some topics.

Tools

Following tools are used in the experiments:

Python and its supporting libraries (Textblob, sklearn, statsmodels) is the primary language scripting in text mining.

Notepadd ++ and Sublime are used for Python scripting

QlikSense Desktop is the Business Intelligence tool for data analytics and visualization

Microsoft Excel for initial data verification and data preprocessing

CHAPTER 1

Introduction

1. Introduction

In this chapter we will cover several areas related to the introduction of the paper. Section 1.1 introduces about the topic of study. Section 1.2 provides an insight towards the important aspects of this research delimitation. Section 1.3 is the introduction of case company. Section 1.4 talks about the formulation of research question. Section 1.5 presents the primary research question along with sub-research questions. Finally, section 1.6 demonstrates the advanced organizer of this paper.

1.1 Topic

Abraham Lincoln once said "*With public sentiment, nothing can fail. Without it, nothing can succeed*." It is always been interesting and an important factor of decision making to know what opinion people carry. In this regard, a sentence or a piece of text that carries sentiment or opinion holds great insights for consumers and organizations. Therefore, information retrieval and information extraction play important roles to locate and extract valuable information out of unstructured data. Opinion mining stays the hub of information retrieval, information extraction and data mining. Many Web 2.0 applications developed after 2004 have also created an abundance of user-generated content from various online social media platforms such as forums, online groups, web blogs, social networking sites, social multimedia sites including photos and videos, and even virtual worlds and social games. According to Chen et al. (2012), in addition to everyday events, and socio-political sentiments expressed in these media, Web 2.0 applications can efficiently gather a large volume of timely feedback and opinions from a diverse customer population for different types of businesses.

This new form of communication has revolutionized how we are interacting with other people, and gradually changed the behavior of consumers on the Internet. The increased availability of information on social platforms has changed the common customer into a researcher. Instead of passively retrieving information from the Internet before making the purchase decision, the customer can now log on to the Internet and review other users' experience about specific products or how a company handles its customer service. The more information available, the more likely the customer will make the buying decision that fits his or her needs. Similarly, he or she will also be encouraged to express their own experience about products and services on social platforms.

The emergence of customer-generated content on various platforms offers great opportunity for researchers and practitioners to listen to the voice of customers, employees and the media. The challenge that comes along is that the data is unstructured but often it contains rich information like opinions and behavioral information about customers. This information is of great value for organizations to consider during market research and product development strategies.

These customer reviews are such informative source of data that organizations can take advantage of them and derive valuable information, which can be further interpreted to customer satisfaction level and product quality assessment. In order to tackle with this context, enterprise systems need to address the need of numerous customers, each having unique perspectives towards the product quality and functionalities.

In order to efficiently address the opinions of countless consumers, system analysts must take into account a wide variety of consumer preferences, not only from existing consumers, but also from potential ones. Organizations can utilize the content, either to do manual analysis of data or perform content analytics using text analytics technology to gain insights into customer sentiment and behavior.

Traditionally, companies can use several methods like interviews or questionnaires so that they can have direct contact with their customers and listen to them. However, these conventional methods explicitly show critical disadvantages in the development of public opinions due to the fact that they are unable to be performed on a large-scale. Although information provided by surveys and interviews is highly tailored for customer opinions discovery purpose, it is not feasible to interview more than a few dozens of customers, or several thousands in case of questionnaires.

The alternative method to collect opinions from public via the Internet has been popular lately. Deep analysis into various customer opinions would provide a solid foundation for companies to evaluate their product quality and create better designs so as to satisfy customers' increasing demands in the future.

Social data has become one of the largest data banks available to companies looking to learn more about their customers. This source of data could reflect the most updated trends and real-time opinions of the public. Moreover, these social platforms provide a low-cost and easily accessible source of data for consumer value analysis. However, it is such a challenge to filter and extract useful piece of information from data due to the fact that the results can be skewed because current technology has trouble picking up on tone, slang and nuances such as sarcasm when trying to interpret text data.

In addition, processing social data and extracting useful piece of information from consumer opinions cannot be done purely on manual basis due to the large amount of data (Zeng et al, 2010). Therefore, the primary objective of this paper is to demonstrate the ability of data analytics in capturing and translating consumer opinions into sentiment identification with the application of automation techniques. By applying various data analytics techniques, we would like to propose an effective model to transform consumer reviews into useful piece of information to support the decision making process in organizations.

The model developed in this dissertation is demonstrated in a case study, where social data related to our primary data point (Jabra products) is retrieved primarily from popular e-commerce websites such as Amazon.com or Bestbuy, and then analyzed by a set of text mining techniques. With the prominence of these websites, we can collect various consumers' opinions about Jabra products, so as to assess and evaluate the quality of their product features.

Thus, by applying text mining techniques, our paper demonstrates the ability of social data that can help companies have an overall evaluation about their current product quality. The result of this experiment can also help the case company with recommended features so as to take into consideration for future product strategy.

1.2 Delimitations

The intention for this dissertation is to analyze the social media customer reviews from Web such as Amazon.com, in order to capture customers' feedback regarding their experience with Jabra's product and product features. We also want to perform sentiment analysis based on the customers' opinions, and make use of the results to help management team at Jabra make informed decisions regarding product quality and development. In order to explain the delimitation or scope of this study, we have divided it under three main subheadings; Importance, Business relevance and Motivation.

1.2.1 Importance

Before the advent of the Internet, it was the premise of Customer Relationship Management tools, that the organizations could manage relationship with their customers to maximize lifetime value. However, this objective only benefits the organizations. Presently, social media and other new technologies have empowered consumers and enabled them to filter out advertising and messages, compare prices with competitors from anywhere with mobile devices, and distribute positive or negative brand messages to a global audience.

As a result, many companies attempt to engage social media in their customer acquisition efforts beginning by uploading advertising spots on YouTube, running promotions on Facebook, or providing information about their products on Wikipedia (Malthouse et al., 2013). These actions can help companies create awareness and change attitudes among prospective customers, thereby, contributing to the acquisition of new customers.

According to Chen et al. (2012), recently data analytics have been used to describe the data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex that they require advanced and unique data storage, management, analysis, and visualization technologies. Chen et al. (2012), further argue that web intelligence, web analytics, and the user-generated content collected through Web 2.0-based social and crowd-sourcing systems (Doan et al. 2011; O'Reilly 2005) have ushered in a new and exciting era of business intelligence and analytics (BI&A) research in the 2000s, centered on text and web analytics for unstructured web contents.

The term business intelligence (BI) became a popular term in the business and IT communities only in the 1990s. However, Davenport (2006) as cited in Chen et al. (2012) posit that it was in the late 2000s, that the term business analytics was introduced to represent the key analytical component in BI.

1.2.2 Business Relevance

According to Popescu & Etzioni (2005), the Web contains a wealth of opinions about products, politicians, and more, which are expressed in newsgroup posts, review sites, and elsewhere. Product reviews on Websites such as Amazon.com often associate metadata with each review indicating how positive (or negative) using a 5-star scale, and also rank products by how they rate in the reviews at the site.

Organizations selling products on the Web motivate their customers to share their opinions and hands-on experiences on products they have purchased. According to Jin et al. (2009), unfortunately, reading through all customer reviews is difficult, especially for popular items, where the number of reviews can be up to hundreds or even thousands. This makes it difficult not only for a potential customer to read them to make an informed decision about the product, but also for the organization to take these product reviews into consideration and use these reviews to improve product quality and customer satisfaction. Jin et al (2009) argue that, to overcome the problem of reading reviews and learning from customer experience, "opinion mining" has seen increasing attention over the last few years.

Analyzing the online user generated content is not only for the marketing practice, as for many organizations it is the key step in design and development of new products and repositioning of existing products. These consumer reviews not only identify the physical product and attributes but also provide a significant insight into the quality of services and support.

Our aim for the analysis of online product reviews is to support managerial decision making in at least two ways. First, the reviews can serve as a filter for other attribute elicitation methods; attributes that experts and also customers identify may have more importance for purposes of product marketing and design. Second, our approach can identify significant attributes that experts otherwise overlook. Due to the customer's ability to write comments as online product reviews, it provide greater insight for the organization as they carry information about what product or attribute customers like or dislike along with the reason as why.

1.2.3 Motivation

Customers' motivation

The low cost of IT technology and availability of the Internet contribute towards the huge amount of information generated every second. The transition of easy to use technology have all the facilities necessary for easy installation, content publication and edition, enabling a common user to publish online content without being a computer expert or a programmer. Due to the availability and ease of use, the number of active users has increased dramatically, resulting in huge amount of content generated by users. Moreover, the concept of Web 2.0 has changed passive users of technology into active participants of online social communities and forums.

Digital Motivation

To meet the new demand of digitization, most existing platforms like social networks, web magazines, newspapers and e-commerce change their systems to comply with new design standards that are introduced by Web 2.0. Among other things, Web 2.0 provides a common space allowing interaction of users through the exchange of opinions and experiences. As a result, huge amount of data has been produced by users, a valuable content that can be extremely useful as primary or complementary source of information. The content in the form of comments, discussion or opinion is unstructured data that needs to go through a process of extraction, filtering and specialized techniques to analyze it and extract meaningful information.

Research Motivation

The huge amount of data available in form of customer generated content needs special attention as how to handle information overload and ensure that the user will have access to the best sources with the least effort. Recently, a lot of attention has been paid to the amount of user-generated content especially the e-commerce sector is the most affected one by the amount of data produced by customers. Customer's opinions represent a valuable unique type of information which should not be ignored by the research community.

Business Motivation

From customer's perspective, mining others' opinions before purchasing a product is a common behavior long before the trend started on the Internet. However, from the business perspective, receiving consumer's feedback can greatly improve its strategies in order to increase profits and customer satisfaction. Usually, websites allow customers to rank their products or write an opinion. However, the product ranking lacks that depth of information as compared to an opinion in the form of text that can portray better reality of the customer experience.

The potential response to consumer reviews includes complaints, product experience and satisfactory or unsatisfactory word of mouth about the product or retailer. Traditionally, organizations have been interested in customer satisfaction, and the reason why customers continue to purchase those products with which they are satisfied (Hallowell, 1996). Customers tell others about particularly pleasing products, and this may influence the brand perceptions of those with whom they communicate. However, the aspect of consumer dissatisfaction needs to be considered in order to improve the product quality and maintain long lasting relationship with satisfied customers. Hallowell (1996) believes that the potential impact of customer reviews can be quite significant on the product success, as the dissatisfied customers would spread the word about their experience.

Dissertations' Motivation

Despite this widespread belief about the importance of online generated content, there is little focus on literature documenting that community content plays important role in consumer decision making. It seems that such a finding is a necessary prerequisite for content provision to be a profitable strategy. Thus, this study emphasizes the need of special mechanisms that aims to provide the organizations with better ways to take full advantage of consumer generated content.

In this dissertation, we would like to examine the effect of consumer reviews on Jabra product development pattern. This paper evaluates the reviews that can be used by Jabra quality management to improve the product quality and comfort of audio device.

Consumers' experience and experience for audio device are investigated through consumer reviews based on Amazon.com and Bestbuy. It is hoped that these reviews and customer experience will provide insights to Jabra management to improve/maintain product quality.

1.3 Case Company Introduction

Jabra is a pioneering innovator of new sound experiences from wired to wireless headsets for mobile users, contact centers and office-based users. Jabra Corporation is a premium brand of GN group. According to Wikipedia, Jabra Corporation has among its roots Norcom Electronics Corporation, a Utah corporation founded in 1983 by inventor/entrepreneur Elwood "Woody" Norris. From its inception, Norcom was engaged in the development of ear-radio and ear-microphone technologies. Jabra Corporation was founded on January 3, 1993. Jabra developed the first in-ear integrated microphone and speaker, invented and patented EarGels, developed DSP based echo and noise cancellation technologies, became the first company to tune a headset over-the-air, and created the market segment category for mobile headsets with its Motorola StarTac product. In 1996, the company acquired operations in Scotland to lead its push into Bluetooth technology. In September 2000, Jabra Corporation was acquired by GN Netcom, a division of the Danish company GN Great Nordic, the latter founded in 1869 as the Great Northern Telegraph Company is listed on NASDAQ Copenhagen.

In 2006, GN Netcom consolidated its Contact Center and Office (CC&O) headset division under the Jabra brand. This was followed by a restructuring in 2008, which established two distinct divisions within Jabra; CC&O and Mobile. This restructuring thus facilitated a greater focus on business-to-business and consumer markets respectively.

1.4 Problem Formulation

This study aims at analyzing consumer opinions about Jabra products aspects that are usually posted and discussed on popular e-commerce websites. Although most retail stores encourage their consumers to write product reviews on their websites, consumer communication is no longer limited to one medium only. Similarly, beyond the Jabra's open platform for user reviews and rating, a number of consumer's reviews can be retrieved from common e-commerce websites. Empowered by the Web 2.0, modern consumers are now easy to share their experience and reviews, reaffirming their active roles and participation in the consumer communication and experience sharing. It also contributes to the transformation of purchasing process, where consumer is now playing the decisive role. This shift has brought companies involving in the process to expose to a new challenge of understanding their consumer expectations, which recently vary on different platforms. By having deep insights into various consumer experiences, the companies can have an exact evaluation of their product quality, and perhaps design better system functionalities in future product release.

Based on explained above insights regarding consumer reviews and ease of use related to social platforms, this study will focus the aspects of social content generated by consumers. We would like to use this data to firstly perform the sentiment analysis about particular product features that customers give comments on social platforms. After that, we would also like to evaluate specific features with their sentimental value, and assess the quality of their product features. We believe that this result would be beneficial for the case company in their product quality assessment.

1.5 Research Question(s)

In the light of the research problem, the following research question is formulated: "*What is the effect of text mining techniques on quality management of Jabra, using consumers' product reviews?*"

The major steps of this process as presented in the research problem proposes sub-research questions to be derived from research question as follows:

- What is the effect of automation techniques using text mining on evaluating consumer sentiment?

- What is the effect of automation techniques on the process of extracting key features from reviews?

Answers to the aforementioned sub-questions solves a portion of the research problem by highlighting the chances for automation within the process, as well as providing analysts with insights to address the primary research question.

1.6 Advanced Organizer/ Dissertation Structure

The structure of this dissertation report is organized as follows.

Chapter 1 provides a brief introduction about the overall background of the dissertation.

In **chapter 2**, we present the conceptual frameworks applied for our experiments to help understand research question and data analytics methods and techniques.

In **chapter 3**, we present the practical relevance of our experiment by explaining related academic articles performed similar research and similar data analytics methods in the past.

Chapter 4 is our explanation for methodology chosen for our dissertation.

In **chapter 5**, we present the results from the experiments.

In **chapter 6**, we discuss the outcome of our experiments in light of conceptual framework.

In **chapter 7**, we answer our research question and present the implications of our study, its limitations and suggestions for further analysis. Finally, we conclude the paper reflecting back on the different processes performed, while offering an evaluative summary with final remarks.

CHAPTER 2

Conceptual framework

2. Conceptual Framework

In this chapter, we would like to explore opinion mining and related concepts in order to understand its importance for the sake of answering our research question. The concepts are focused on e-commerce; however, the same principles can be used in many different domains that are likely to admit opinions. The concepts and techniques related to data analytics are also discussed below in this chapter.

2.1 Concepts that help understand the research question

2.1.1 Social media and two way communication

Kaplan and Haenlein (2010), describe that the era of social media probably started about 20 years ago, however, the accessibility and high speed of the Internet has immensely increased the popularity of social networking sites. This led to the term "social media" and contributes to the importance that it has today (Kaplan and Haenlein, 2010). Authors explain that Web 2.0 is considered as the platform for the evolution of social media and user generated content and can be seen as the sum of all the ways in which people use social media.

According to Organization for Economic Operations and Development (OECD, 2007) (as cited in Kaplan and Haenlein 2010) the user generated content needs to fulfill three basic requirements in order to be considered as such; First of all, it needs to be published either on a publicly accessible website or on a social networking site, accessible to a selected group of people; second, it is important that the content shows a certain amount of creative effort; and last but not the least, it must be created outside of professional routines and practices. Based on the knowledge about Web 2.0 and user generated content, a clearer and more precise definition of social media was derived by Kaplan and Haenlein (2010). "*Social media is a group of Internet based applications that build on the ideological and technological foundations of Web 2.0, and that allow the exchange and creation of user generated content*".

The principle of social media branding relies on the word of mouth idea, where online media users speak about e.g. brands and products. These conversations could help to differentiate and understand the brand. The affordance of social media includes features and the themes of the content, which is "*shared and spoke about*" by the audience (Kaplan & Haenlein, 2010).

The unique characteristics of social media have empowered users with active role and participation on the online platforms resulting in changing the process of market communication. Ketter and Avraham (2012), explain that before the advent of social media's innovative marketing techniques, the marketing campaigns and models were based on the linear model of communication with a top-down flow of information. The techniques focused on sources e.g. T.V or Radio, to deliver a well-chosen message to a rather passive audience, using selective channels and aiming for a specific consequence (Ketter and Avraham 2012). However, the role of the user has shifted as active producer of content after the digital shift of Web 2.0, where a new approach began and consumer also act as active sources that can create and distribute product reviews and experience messages.

2.1.2 Word of mouth & User generated content

According to Chevalier & Mayzlin (2006), online user reviews have become an important source of information to consumers, substituting and complementing other forms of business-to-consumer and offline word-of-mouth communication about product quality. Consequently, many managers believe that a Website must provide community content to build brand loyalty (Fingar, Kumar, and Sharma 2000; McWilliam 2000 as cited in Chevalier Mayzlin 2006).

Lange-Faria & Elliot (2012), define social media by the idea of allowing many users of the Internet to access, share, collaborate and update web content. The definition is rooted in community; users may engage, collaborate and share with others in real time (in the case of virtual and mobile technologies) without constraint of time or geography. The authors argue that user generated content is quickly becoming the source for credible information with social media becoming the primary medium by which information is shared. Search engine results particularly show social media sites and opinions of people in the form of user generated content is primary vehicle by which the consumers judge the products (O'Connor, P., Wang, Y., & Li, X. 2011 as cited in Lange-faria & Elliot 2012).

According to Li and Hitt (2008), the term word-of-mouth has long been recognized as a major driver of product sales, as it not only increases consumer awareness, but it may also be one of the only reliable sources of information about the quality of experience goods (i.e., products not easily characterized prior to consumption). With the development of the Internet, word-of-mouth has moved beyond small groups and communities to being freely available through large-scale consumer networks (Avery et al. 1999 as cited in Li & Hitt 2008). These networks have magnified the depth and span of word-of-mouth to an unprecedented scale. Online opinion and consumer review sites have correspondingly changed the way consumers shop, enhancing the traditional sources of consumer information such as advertising. Li & Hitt (2008) support their argument that a survey of 5,500 Web consumers conducted by BizRate, 44% of respondents said they had consulted opinion sites before making a purchase and 59% considered consumer-generated reviews more valuable than expert reviews. In some product categories such as electronics, surveys suggest that online review sites have a greater influence on product purchase decisions than any other sources (DoubleClick 2004 as cited. in Li & Hitt 2008).

As mentioned by Dellarocas (2006), one reason why consumer generated reviews may not represent actual product quality is due to "forum manipulation," in which firms hire professional reviewers or may encourage friends and colleagues to artificially boost the ratings of their products. His results suggest that even in the presence of manipulation, reviews are still informative because producers of the highest-quality products also receive the greatest benefit from manipulation.

2.1.3 Credibility of online customer reviews

In order to study the effect of online customer reviews Mudambi & Schuff (2010), explain that customer reviews are increasingly available online for a wide range of products and services. These reviews supplement other information provided by electronic storefronts such as product descriptions, reviews from experts, and personalized advice generated by automated recommendation systems. These reviews play a vital role in consumer decision making, so it is important to understand what makes these reviews credible information both for consumers to base their purchase decision and the organization, in order to base their product development and marketing decisions (Mudambi & Schuff 2010).

Authors claim that the review depth has a greater positive effect on the helpfulness of the review for search of products than for service goods. Mudambi & Schuff (2010), explain that online customer reviews can be defined as peer-generated product evaluations posted on company or third party websites. Retail websites also offer consumers the opportunity to post product reviews with content in the form of numerical star ratings (usually ranging from 1 to 5 stars) and open-ended customer-authored comments about the product. Leading online retailers such as Amazon.com have enabled consumers to submit product reviews for many years. Mudambi & Schuff (2010), further add that some other firms choose to buy customer reviews from Amazon.com or other Websites and post the reviews on their own electronic storefronts. In this way, the reviews themselves provide an additional revenue stream for Amazon and other online retailers.

Kumar & Benbasat (2006) posit that the bond between company and its customers can be examined along two distinct relationship orientations—transactional and social (Clark et al. 1986, Mathwick 2002 as cited in Kumar & Benbasat 2006). The transactional orientation focuses on the utilitarian nature of the relationship between a customer and the company (e.g., usefulness), and the social focuses on the relational nature that is exemplified by feelings of intimacy and warmth (e.g., social presence). Research has suggested that a website's support for personalization is a key feature that can play a significant role in enhancing its customers' shopping experience. Kumar & Benbasat (2006) explain the term personalization in the context of online shopping as providing information and applications that are matched to the interests, roles, and needs of a visitor to a website.

2.1.4 Perceived Quality

Tsiotsou (2006), define perceived quality as the consumer's judgment about a product's overall excellence and superiority. According to Zeithaml (1988), perceived quality differs from objective quality. Perceived product quality is a global assessment characterized by a high abstraction level and refers to a specific consumption setting. The term "objective quality" is related closely to the concepts used to describe technical superiority of a product. As Tsiotsou (2006) describes that the objective quality refers to the actual technical excellence of the product that can be verified and measured. However, Zeithaml (1988), argues that objective quality may not exist because all quality is perceived by someone, who can be a consumer or manager or researcher. However, the concept of quality might be perceived differently as what is important to managers might be different from consumer's view. Zeithaml (1988), describes that a research study for General Electric, points out striking differences between consumer and manager perceptions of appliance quality. In his paper, Zeithaml (1988), explains that when managers were asked how consumers perceive quality, managers listed workmanship, performance, and form as critical components. However, consumers actually keyed in on different components: appearance, cleanability and durability.

Olshavsky (1985), defines the ability to infer quality from specific attributes as "surrogate-based preference forming behavior" and cites examples of product categories in which a given surrogate is highly associated with quality, for example, size signals quality in stereo speakers or style signals quality in cars and clothes. Olshavsky (1985) explains in his paper that in the consumer perceived quality study, consumers repeatedly associated quality in fruit juices with purity as 100% fruit juice with no sugar added or freshness. In these examples, one or a few attributes from the total set of attributes appear to serve as reliable indicators of product quality. Olshavsky (1985), admits that generalizing about quality across products has been difficult for managers and researchers, as attributes that define quality in fruit juice are not the same as those indicating quality in washing machines or automobiles. Even within a product category, specific attributes may provide different signals about quality. Zeithaml (1988), argues that research shows that there are several indicators of perceived quality, as long term advertising appears to function as an alternate for quality when the consumer has inadequate information about intrinsic attributes. Brand names also serve as "shorthand" for perceived quality products.

Zeithaml (1988) points that consumers' perceived quality depending on different stages of product purchase process. The perceived quality of the product before purchase; at time of product involvement and purchase intention, might be different than after purchase and user experience also called customer satisfaction.

2.1.5 Customer satisfaction

According to Tsiotsou (2006), customer's satisfaction has been considered one of the most important constructs, and one of the main goals in marketing (Morgan et al., 1996; McQuitty et al., 2000 & Erevelles and Leavitt, 1992). Satisfaction plays a central role in marketing because it is a good predictor of purchase behaviour that includes repurchase, purchase intentions, brand choice and switching products behaviour (McQuitty et al., 2000 as cited in Tsiotsou, 2006). Due to the topic's importance, various theories and models have been developed in an effort to define the construct and explain satisfaction during different products/services and consumption stages. Tsiotsou (2006) claims that satisfaction is an important predictor of customer loyalty leading towards strong customer relationship. The strength of the relationship between the customer and product developer is strongly influenced by customer characteristics such as variety seeking, income and demographic variables like education and age. Tsiotsou (2006) explains that satisfied customers tend to use a service more often than those not satisfied as they present a stronger repurchase intention, and they recommend the service to their acquaintances.

2.1.6 Automated market research

According to Lee & Bradlow (2011) marketing research, the set of methods to collect and draw inferences from market-level customer and business information, has been the lifeblood of the field of marketing practice and the focus of organizations. Lee & Bradlow (2011), argue that from practical perspective, this has brought forward the “stalwarts and toolbox” of the marketing researcher, including methods such as preference data collection using conjoint analysis (Green and Srinivasan 1978 cited in Lee & Bradlow 2011) inferring market structure through multidimensional scaling (Elrod 1988, 1991; Elrod et al. 2002 cited in Lee & Bradlow 2011), inferring market segments through clustering routines (DeSarbo, Howard, and Jedidi 1991 cited in Lee & Bradlow 2011), or simply understanding the sentiment and “voice of the consumer” (VOC; Griffin and Hauser 1993 cited in Lee & Bradlow 2011). Although these methods are here to stay, the radical changes resulting from the Internet and user-generated media promise to fundamentally alter the data and collection methods used to perform these methods.

Lee & Bradlow (2011), describe that automated market research is a text-mining algorithm for analyzing online customer reviews to facilitate the analysis of market structure in two ways. First, the voice of consumer, as presented in user-generated comments, provides a simple, principled approach to generating and selecting product attributes for market structure analysis. Second, the bulk of opinion, as represented in the continuous stream of reviews over time, provides practical grounds to shift from traditional approaches such as surveys and focus groups for conducting brand sentiment analysis and can be done continuously, automatically, inexpensively, and in real time (Lee & Bradlow 2011).

2.2 Concepts that inform about the data analytics methods and techniques

2.2.1 Sentiment analysis or opinion mining

Richins (1983) emphasizes that the opinion mining tool aims to mine customer reviews of a product or attribute and extract high detailed product entities on which reviewers express their opinions. Opinion expressions are identified and opinion orientations for each recognized product entity are classified as positive or negative. The impact of user generated content opinion mining is great from consumer perspective as well as business perspective. The recommendations and consumer reviews on customers' products provide significant information about customer satisfaction and product performance in the market. Organizations can leverage great benefit by taking these reviews into consideration while tracking business performance.

The basic task of opinion mining is polarity classification. Polarity classification occurs when a piece of text stating an opinion on a single issue is classified as one of two opposing sentiments. Equivalently, sentiment analysis is the process of identifying the polarity of opinion in a given stream of text (Pang et al. 2008). It uses natural language processing, text analytics and computational techniques to automate the extraction or classification of sentiment from sentiment reviews. The opinions, emotions and evaluations captured in the text is often classified to distinct labels named positive, neutral or negative (Wilson et al, 2005). The primary goal of sentiment analysis is to analyze the reviews and examine the scores of sentiments (Hussein, 2016). Pang & Lee (2004) claim that the purpose of sentiment analysis is to seek the viewpoint that is underlying a text span; for example, extracting the meaning behind "thumbs up" or "like" to determine the sentiment. They further explain that the idea of "what other people think" has always been an important piece of information for most of us during decision-making process.

There exist multiple approaches to sentiment analysis (Cambria et al, 2012). For example word spotting, which is based on the presence of unambiguous words like happy, sad, great and awful; or statistical methods, which calculates the polarity of keywords and co-occurrence frequency based on a large training corpus. Depending on the level at which the analysis is made in the text, sentiment analysis techniques can be divided into three categories: document level, sentence level and entity/aspect level (Rodrigues, 2016). According to Hussein (2016), the highest technique usage in the theoretical type of sentiment analysis is parts-of-speech (POS) tagging and lexicon-based techniques, while n-gram technique is the highest technique used in the technical type.

2.2.2 Content Analysis

Krippendorff (2004), posits that the term content analysis is about 60 years old. According to the author, Webster's Dictionary of the English Language included the term in its 1961 edition, defining it as *"analysis of the manifest and latent content of a body of material such as a book or film through classification, tabulation, and evaluation of its key symbols and themes in order to ascertain its meaning and probable effect*". The intellectual roots of content analysis, however, can be traced far back in human history, to the beginning of the conscious use of symbols and voice, especially in writing.

According to Krippendorff (2004) content analysis is a research technique for making replicable and valid inferences from texts to the contexts of their use. During 1950s researchers witnessed considerable interest in mechanical translation, mechanical abstraction, and information retrieval systems. The large volumes of written documents to be processed in content analysis and the repetitiveness of the coding involved, made the computer a natural but also a difficult ally of the content analyst. The development of software for data processing stimulated new areas of exploration, such as information retrieval, information systems, computational stylistics (Sedelow & Sedelow, 1966 as cited in Krippendorff. 2004), computational linguistics, word processing technology, and computational content analysis.

Content analysis has evolved into a repertoire of methods of research that promises to yield inferences from all kinds of verbal, pictorial, symbolic, and communication data. Krippendorff (2004) proposes a conceptual framework for content analysis within which that role becomes clear. This framework is intended to serve three purposes: Its prescriptive purpose is to guide the conceptualization and design of practical content analytic research; its analytical purpose is to facilitate the critical examination and comparison of the published content analyses; and its methodological purpose is to point to performance criteria and precautionary standards that researchers can apply in evaluating ongoing content analyses.

2.2.3 Text analytics and Natural language processing (NLP)

Text analytics refers to the process of deriving high-quality information from text. Text analytics works by transforming words and phrases from unstructured data into numerical values which, then can be further processed by traditional mining techniques. NLP is an area of research which examines the ability of computers in understanding and manipulating natural human languages. NLP is closely linked to other fields of science, such as linguistics, cognitive science, psychology and philosophy. NLP is proved to have a significant number of applications and among them; customer experience management is a big application for retail industry. E-commerce companies enjoy a large base of customers who increasingly express their needs, attitudes, preferences and frustrations online. Social platform listening has become an important tool for e-retailers who want to understand consumer shopping habits, predict product demand or monitor trends to create sticky marketing messages.

2.2.4 Business Intelligence

According to Golfarelli et al., (2004) business intelligence can be defined as the process of turning data into information and then into knowledge. The knowledge gained from business intelligence is usually about customer needs, customers decision making processes, industrial environment, and general economic, technological, and cultural trends. Negash (2004) defined a business intelligence system as a combination of data gathering, data storage, and knowledge management with analytical tools, in order to support planners and decision makers with complex internal and competitive information. This means that a good business intelligence system would be able to facilitate management team with actionable information at the right time and right place.

2.2.5 Term Frequency - Inversed Document Frequency (TF-IDF)

The TF-IDF is one of the most common methods that data analysts use to evaluate the importance of words/phrases in regards to the document within a corpus. It is a numerical statistic value that reflects how important a word is towards a document in the context of a corpus. It is usually used as a weighting factor in information retrieval and text mining.

The TF-IDF is in fact a production of two statistics, term frequency and inversed document frequency, which have various methods to calculate. Term frequency refers to how frequent a term appears in the sentence, while the inversed document frequency is a measure of how much information the word provides, that is, whether the term is common or rare across all documents. The TF-IDF value of a corresponding term must therefore be higher when this term occurs frequently in a particular document, but it must also occur rarely in the whole set of documents examined.

According to Chen (2016), TF-IDF has a disadvantage of creating a bias that frequent terms highly related to a specific domain are typically identified as noise, thus leading to the development of lower term weights. It is due to the fact that the traditional TF-IDF method is not specifically designed to address large corpus. However, this bias can be overcome, provided that the class labels are prepared in advance. Considering our dataset which contains approximately 3,000 reviews with predefined labels, we can use TF-IDF as an appropriate method for term weighting.

2.2.6 Part-of-speech (POS) tagging

According to Mitkov (2005), Part-Of-Speech, or POS tagging is, as the name implies, the process of assigning a part-of-speech, such as noun, verb, adjective, etc., to each word in a sentence. For the purpose of this study we applied part of speech information to deduce the opinion and subjective information from the given text. Adjective and verbs play an important role in deducing the subjective information since they reflect the qualitative judgment about a text. After tagging part-of-speech we apply our algorithm and based on the score we get from both the positive and negative model, we deduce the nature of the opinion.

2.2.7 Ordinary Least Squared (OLS) Regression Analysis

Dismuke & Lindrooth (2006), mention that OLS regression analysis is one of the most common techniques used in multivariate analysis. OLS regression is defined as a generalized linear modelling technique that may be used to model a single response variable which has been recorded on at least an interval scale (Moutinho & Hutcheson, 2011). It is a method that helps analysts estimate the unknown parameters in a linear regression model. It has the primary goal of minimizing the sum of squares of differences between observed responses and those predicted by a linear function of a set of explanatory variables. The technique may be applied to single or multiple explanatory variables and also categorical explanatory variables that have been appropriately coded. OLS regression is particularly powerful as it relatively easy to also check the model assumption using simple graphical methods (Hutcheson & Sofroniou, 1999).

In an OLS regression analysis, there are dependent and independent variables. A dependent variable is denoted as Y in a regression model, and it is presented on the left-hand side. The independent variables are usually a series of X_i*s.* The general expression of OLS regression is as follows:

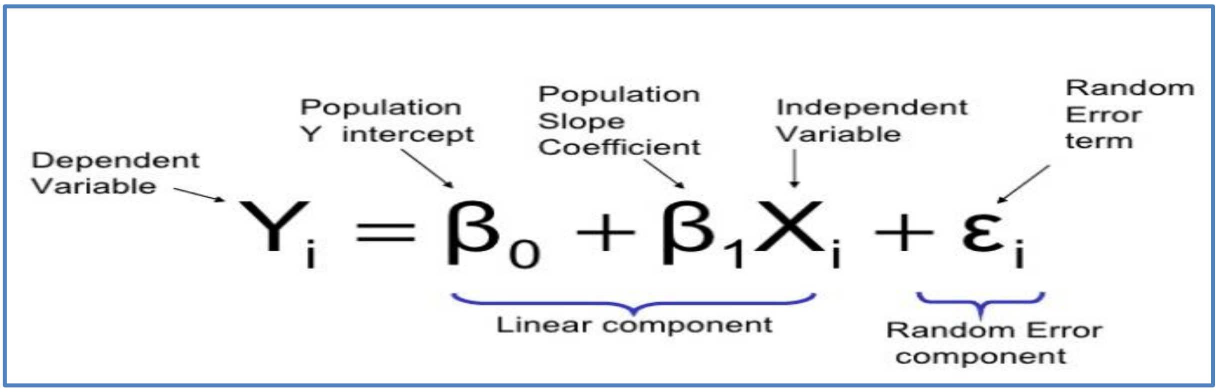


Figure 1 - General expression of OLS regression

In relation to the result of OLS regression analysis, there is a factor that needs to be carefully interpreted, which is the coefficient. It is in fact a measure of the linear correlation between two variables X and Y. It has a value between +1 and −1, where 1 is total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation. This factor helps analysts identify the correlation between the explanatory variables and the predicted variables.

2.2.8 Data visualization

Data visualization for example, charts, graphs or informatics is an effective method to communicate important information to audience at a first glance. For text-based data, the most commonly used method for data visualization is word cloud. Word cloud (or tag cloud) is used for visually summarizing large amounts of text aesthetically using various font sizes and color of words to map with the word frequency, popularity or importance (Wang et al., 2014). It has received high attention from both researchers and industry, and widely used in both business and research, like Wordle, TagItOut or TIRRA (Liu et al., 2012)

CHAPTER 3

Previous work

3. Previous Work

In this chapter we would like to briefly discuss the concepts and theories used in our dissertation, in light of earlier research.

Our work is closely related to Dave, Lawrence & Pennock's (2003), Li & Hitt (2008), Morinaga et al.(2002), Liu & Hu (2004) and similar works on semantic classification of reviews. Using available training corpus from some Websites, where each review already has a class e.g., thumbs-up and thumbs-downs, or some other quantitative or binary ratings, the above mentioned researchers designed and experimented a number of methods for building sentiment classifiers. They also used their classifiers to classify sentences obtained from Web search results, which are obtained by a search engine.

Our study is based on Liu & Hu (2004); we are interested in features of the product that customers have opinions on and also whether the opinions are positive or negative. We do not summarize the reviews by selecting or rewriting a subset of the original sentences from the reviews to capture their main points as in traditional text summarization. We make use of both data mining and natural language processing techniques to perform this task and also present a comparative evaluation based on Liu & Hu (2004).

3.1 Academic Articles on the Case Company's Industry Sector

The research in the field of data mining and sentiment analysis is quite vast. Nonetheless, the research in the area of customer opinions about headphone and headset industry is still untouched as it is very difficult to find work focusing on the area of opinion mining about headphone/headset industry. It is important to analyze the functionality of headphones/headsets in order to develop an understanding of attributes and the customer quality assessment of the product features. Therefore, we would like to have an overview of headphones description and product feature claims related to it.

Lee et al. (2009)

According to Lee et al. (2009) portable entertainment and communication equipment have been proliferating, including devices such as cellular phones, MP3 players, and portable computing devices. In all of these examples, audio communication is a large part of the user experience. It is often desirable to use headphones when listening to music and other audio material. In order to increase convenience and audio quality and to provide privacy, one-way headphones or two-way headsets are employed. For added convenience, wireless headphones/headsets are available. For example, Bluetooth headsets are available for telephone conversations as well as headphones for audio listening.

Mozer et al. (2012)

Mozer et al. (2012) describe that there is a need for improving user interaction with electronic devices with the continuous development of large scale integration electronics and data storage technologies. Consumer electronic products are getting smaller and smaller. However, their computational capabilities are growing such that they are able to perform more and more functions in a smaller form factor. Therefore, the authors explain that the smaller form factor and lighter weight allow the creation of nearly invisible electronic products that fits on a person's ear or onto their clothing. These developments make it more difficult for users to interact with electronic products because the number of buttons required in such smaller footprint products keeps growing due to the trend in increasing complexity and compatibility with cellular phone like the iPhone, to activate and address the increasing feature set. Mozer et al. (2012) argue that the device is small, and there is a limit to the number of buttons on it and there is no conventional keyboard with letters and numbers, making it even more difficult for the user to control the interaction. This forces the user to look at the headset in order to understand the button sequences to know what commands to execute. Authors further explain that forcing the user to refer to the headset somewhat defeats the purpose of the headset since it is desired that the headset sits comfortably and accessible on the person's head, and it can be left nearby (for example in a purse, car seat, briefcase, purse, trunk, packet etc.), but does not need to be removed frequently.

Trompette & Chatillon (2012)

Trompette & Chatillon (2012) studied the call center background and analyzed that noise does not contribute to noise exposure; it impacts working conditions and influences the headset volume setting. It was therefore measured at the same time as exposure to noise. Their study reveals that the risk of hearing impairment was generally low.

Besides exposure to noise, background noise levels are often high with regard to recommendations for office workers. Their application is intended to ensure the absence of excessive exposure to noise and improve acoustic comfort.

3.2 Academic Articles on Similar Research Questions

We note that our study is certainly not the first to employ user-generated content or even specifically online reviews for the purposes of product quality assessment. The impact of customer reviews on consumer behavior has long been a source of study. A large body of work has explored how reviews reflect or shape a seller's reputation (Dellarocas 2003; Eliashberg and Shugan 1997; Ghose and Ipeirotis 2006). Other researchers have studied the implications of customer reviews for marketing strategy (Chen and Xie 2004). That said there has been comparatively little work on what organizations might learn from customer reviews for purposes of studying market structure and customer preferences. Our objective is to learn the full range of product attributes and attribute dimensions voiced in online customer reviews and to reflect that how organizations can use the information to improve product quality, customer satisfaction and eventually increase sales.

3.2.1 Word of mouth and online reviews

Filieri (2015)

Filieri (2015) research adopted dual-process theory (Deutsch & Gerard, 1955) by extending its application to the online environment and investigated the determinants of information identification from the consumer perspective. Moreover, they want to extend the investigation in order to identify what leads to information adoption with regard to online consumer reviews. From the theoretical perspective, the tested model proved that dual process theory can be used to investigate the antecedents of information identification and adoption in the online environment. The findings reveal that the informational influence of e-WOM is stronger than the normative influence and that information quality represents the most important aspect of information identification in e-WOM. This study also found that the normative cues, namely, overall product rankings and customer ratings, show a significant and positive relationship with information identification. However, source credibility was found to exhibit a weak but significant relationship. Information quantity does not appear to influence the dependent variable. This finding implies that reviews with the highest level of information quality and crowd opinions are perceived by consumers to be the most helpful information when becoming familiarized with a product and assessing its quality and performance. Filieri (2015) suggests that in e-WOM, both informational and normative influences play an important role in determining how consumers assess the quality of products.

Li & Hitt (2008)

According to Li & Hitt (2008), even before the emergence of large-scale online communication networks, word-of-mouth was perceived as an important driver of product sales (Rogers 1962; as cited in Li & Hitt 2008). Most of these studies focused on the diffusion of positive experience, which is more related to raising consumer awareness than it is to conveying quality information. The emergence of large-scale online communication networks for the exchange of quality information has led to an emerging literature on the economics of these systems. Considerable research has focused on performance and design of eBay-like reputation systems. However, research on product review systems has been more limited.

The observation that people tend to follow the decisions of others has been extensively discussed in the herding literature, which has attributed this behavior to network externalities (Katz & Shapiro 1985), social sanctioning of deviants (Akerlof 1980), and taste for conformity (Becker 1991). Li & Hitt (2008) work is more closely related to information-motivated herding literature (Banerjee 1992, Bikhchandani et al. 1992) because it is the quality information indicated by early buyers' reviews or ratings that drives later buyers to follow. However, in the cited models, buyers share similar quality perceptions, so herding is the result of rational behavior. (Li & Hitt 2008).

Chevalier and Mayzlin (2006)

Chevalier and Mayzlin (2006) examine the effect of consumer reviews on relative sales of books at Amazon.com and Barnesandnoble.com. The authors demonstrate that the differences between consumer reviews posted on Barnesandnoble and those posted on Amazon.com were positively related to the differences in book sales via the two websites. They further explain that an improvement in reviews leads to an increase in relative sales at that site. However, the impact of one-star reviews is greater than the impact of five-star reviews; and evidence from review-length data suggests that customers read review text rather than relying only on summary statistics.

Chevalier and Mayzlin (2006) characterize patterns of reviewer behavior and examine the effect of consumer reviews on firms' sales patterns. Their findings suggest that, on average, reviews tend to be positive, especially at bn.com. The addition of new, favorable reviews at one site results in an increase in the sales of a book at that site relative to the other site. Evidence shows that an incremental negative review is more powerful in decreasing book sales than an incremental positive review in increasing sales.

Chen and Wu (2004)

Chen and Wu (2004) suggest the mediation role of product recommendations in affecting the relationship between reviews and sales on Amazon.com. Although these studies have established a link between sales and product reviews, they do not examine whether consumer reviews were effective in communicating actual product quality. Moreover, these studies utilize the time series dimension of the data to increase the sample size but do not directly address the time structure of reviews.

Godes and Mayzlin (2004)

According to Godes & Mayzlin (2004), managers are very interested in word-of-mouth communication because they believe that a product's success is related to the word of mouth that it generates. However, there are at least three significant challenges associated with measuring word of mouth. First, related to data collection. Because in case the information is exchanged in private conversations, direct observation traditionally has been difficult. Second, the aspect of conversations that needs to be measured. The third challenge comes from the fact that word of mouth is not exogenous. The mapping from word of mouth to future sales is of great interest to managers, but we must also recognize that word of mouth is an outcome of past sales.

The authors investigate two distinct dimensions of word of mouth: volume and dispersion. These measures are important features due to the fact that these are implementable by the firm at low cost and effort. The first dimension of word of mouth is its volume: How much word of mouth is there? The more conversations there are about a topic, the more people will become informed about it. Godes & Mayzlin (2004), expect that higher volumes of word of mouth will be associated with higher dispersion (marketing), as more informed people are in the community, the more likely they spread the information around.

Zhu & Zhang (2006)

Zhu & Zhang (2006), argue that the new ways of communication pose both challenges and opportunities to the firms. On one hand, it is intimidating to manage the reputation profile generated from the word of mouth process of thousands of people who are complete strangers to each other. On the other hand, the Internet provides the companies with a way to record, retrieve, and measure the reputation profiles. Understanding how consumers react to online reviews is of vital importance to firms who rely on word of mouth to disseminate information about their products.

Their results suggest that online reviews complement offline word of mouth. It is therefore important for firms to design effective online marketing strategies, especially for their less popular products. For example, firms could use promotional chats or invite users to review their products to increase the awareness of their products.

3.2.2 Customer satisfaction & perceived quality

Richins (1983)

According to Richins (1983), number of these studies discussed the appropriate ways to measure satisfaction levels and some addressed the theoretical bases of satisfaction. Firms have traditionally been interested in customer satisfaction, and there is a good reason behind that as customers continue to purchase those products with which they are satisfied. Customer may influence the brand perception by telling others about particularly pleasing products, with whom they communicate. While work progressed in this area, however, less attention was given to consumers' reactions to dissatisfaction. Potential responses include (a) switching brands or refusing to re-patronize the offending store, (b) making a complaint to the seller or to a third party, and telling others about the unsatisfactory product or retailer.

Tsiotsou (2006)

Tsiotsou (2006) explains that marketing managers are interested in consumer purchase intentions in order to forecast sales of existing and/or new products and services. Purchase intentions data can assist managers in their marketing decisions related to product demand (new and existing products), market segmentation and promotional strategies. Tsiotsou (2006) proposes a purchase intention based model that includes deep insights into customer involvement, value, perceived quality and satisfaction that leads to purchase intention. Involvement with the product was positively related to perceived quality, overall satisfaction. Perceived product quality had a direct positive effect on purchase intentions, and it was an antecedent of consumer overall satisfaction. This research provides an improved understanding of the role of these variables on purchase intentions.

Zeithaml (1988)

According to Zeithaml (1988), consumer perceptions of price, quality, and value are considered significant determinants of shopping behavior and product choice. It is also important to differentiate between quality and value from consumer's perspective. Perceived quality is defined as the consumer's judgment about the superiority or excellence of a product. Zeithaml (1988) explains consumers' perceptions of quality change over time as a result of added information, increased competition in a product category, and changing expectations. The dynamic nature of quality suggests that marketers must track perceptions over time and align product and promotion strategies with these changing views. Due to the change in products and perceptions, marketers may be able to educate consumers on ways to evaluate quality. He further advises that, advertising the information provided in packaging, and visible cues associated with products can be managed to evoke desired quality perceptions.

3.2.3 User generated Content

Ketter & Avraham (2012)

The unique characteristics of social media have empowered users with active role and participation on the online platforms, changing the process of market communication. As Ketter and Avraham (2012), explain that before the advent of social media's innovative marketing techniques, the marketing campaigns and models were based on the linear model of communication. As linear model of communication have a top-down flow of information, in which the management choose to deliver a well-chosen message to rather passive audience, using selective channels and aiming for a specific consequence. However, the role of the user has shifted as active producer of content after the digital shift. A new approach began by social production of content where consumer also acts as active sources that can create and distribute campaign messages.

3.3 Academic Articles on Similar Data Analytics

3.3.1 Opinion Mining

Liu & Hu (2004)

Liu & Hu (2004) conducted a study about online customer reviews for popular products. They argued that the huge amount of reviews makes it difficult for potential customers to read them and make informed decision. Liu & Hu (2004), perform classic text summarization using three steps. The first step is mining product features that have been commented on by customers, in the next step they identify opinion sentences in each review and decide whether each opinion sentence is positive or negative and in the last step they summarize and evaluate the results.

The authors proposed a set of techniques for mining and summarizing product reviews based on data mining and natural language processing methods. The objective is to provide a feature-based summary of a large number of customer reviews of a product sold online.

Dini & Mazzini (2002)

Dini & Mazzini (2002) publish a paper studying customer opinions based on search engine results. They apply syntactic and semantic processing to these in order to provide a structured input from natural text, for later processing they used data mining algorithms. Their approach largely consists of shallow syntactic parsing with a method known as chunking, coupled with a semantic parsing algorithm utilizing a "filled template" system to apply a sentiment polarity to sentences part of speech.

Moghaddam & Ester (2012)

According to Moghaddam & Ester (2012), opinion mining is helpful for organizations to analyze customer opinions on their products and features. While product attributes are clearly mentioned, discovering the primary cause behind low profit needs much focus on all the individual customer reviews on such characteristics. Opinion mining is an amazing method for taking care of numerous business trends identified with deals administration, status management, and advertising. Additionally, organizations may have the capacity to perform pattern prediction by following customer perspectives.

Pang et al. (2008)

Pang et al. (2008), focus on methods that seek to address the new challenges raised by sentiment-aware applications, as compared to those that are already present in more traditional fact-based analysis. Authors claim that the Internet and the Web have now made it possible to find out about the opinions and experiences of those in the vast pool of people. These reviewers are neither our personal acquaintances nor well-known professional critics, but only people we have never heard of. Still, more and more people are making their opinions available to strangers via the Internet. With the significant increase of opinion sharing and online customer reviews, there is a greater demand to build better customer shared information access system.

Pang et al. (2008) explain that companies are increasingly coming to realize that consumer voices can yield enormous influence in shaping the opinions of other consumers and, ultimately, their brand loyalties and purchase decisions. Companies can respond to the consumer insights generated through social media and analyze it to help modify their product development, and other activities accordingly.

Morinaga et al. (2002)

Morinaga et al. (2002), proposed a framework about opinion mining, based on automatic opinion labeling in order to assess an entity's reputation. The system was built around extracting characteristic words using stochastic complexity, in order to gain a sense of overall features. These features consists of words generally found to be co-occurring with the feature words are added to the feature for example "display", "text", "email". A set of user specified categories are then tagged with meta-scores indicating how likely an opinion is to be expressed with a typical sentence. This is done via Bayesian theory. Finally, the extracted features are mapped to the opinions of the specified categories with the help of principal component analysis.

3.3.2 OLS Regression Analysis

Schwartz et. al (2013)

In their experiments, social media data is the key data point, which was analyzed by researchers to identify the relationship between words used on social media with personality, gender and age. OLS regression is used in experiment to figure out the correlation between each language feature and each demographic or psychometric outcome. Researchers use OLS regression to link word categories with author attributes, fitting a linear function between independent variables and dependent variables (such as a trait of personality, e.g. extraversion). The coefficient of the target dependent variable is considered as the strength of relationship. Finally, based on the coefficient values, researchers can visualize the correlation between languages on social media and psychosocial variables.

Hallowell (1996)

Hallowell (1996) conducts an experiment to demonstrate the relationship among customer satisfaction, customer loyalty and profitability. In his research, OLS regression is used to examine his hypothesized relationships among these factors. The researcher uses OLS regression to link measures of customer satisfaction with the customer retention, and between customer retention with organization profitability. The coefficient estimates are used to measure the strength of relationships. From the experiment, Hallowell (1996) describes the relationships among three factors, and recommends specific actions for management to optimize their investment.

Christer & Eivind (2012)

In order to understand the relevance of nationality and some other psychosocial variables to the tourists' length of stay in Norway, Christer & Eivind (2012), have deployed OLS regression analysis. In their experiment, they determine the dependent variable as the length of stay, as they wish to explore how it is explainable by other factors. They aim to analyze the stay of length of international tourists, who come to visit Norway during the summer vacation, based on the differences in nationality, types of tours, or level of satisfactions. By applying OLS regression analysis on their collected data, researchers could draw a remarkable conclusion about the impacts of explanatory factors on the tourists' length of stay in Norway during summer holiday. From the result, researchers could suggest profound economic consequences, and recommendations for organizations promoting visitor travelling in Norway.

CHAPTER 4

Methodology

4. Methodology

This chapter presents the choice of research methodology for the dissertation study. The subsections of this chapter are organized as follows: Section 4.1 provides clarification for the choice of research method. Section 4.2 gives a brief introduction about the case company. Section 4.3 describes the dataset used in this research, consisting of detailed description for each fields in the dataset. Section 4.4 summarizes the data analysis process, providing with a data process diagram visualizing all steps taken in the experiments. Accordingly, each section from 4.5 to 4.7 explains processes of data collection, preprocessing and analysis. Section 4.8 introduces some limitations in the dataset, while section 4.9 suggests alternative methods for the research.

4.1 Choice of research method

The purpose of research is to discover answers to questions through the application of scientific research methods. According to Kothari (2004), the aim of research is to find out the truth which is hidden and which has not been discovered as yet. The author explains that although each research study has its own specific purpose, we may think of research objectives as falling into a number of broad categories. One objective could be to gain familiarity with a phenomenon or to achieve new insights into it, and studies with this object in view is termed as exploratory research studies. Another research objective could be to test a hypothesis of a causal relationship between variables and this research is known as hypothesis-testing research studies.

The above description provided by Kothari (2004), about the types of research brings to light the fact that there are two basic approaches to research; quantitative approach and the qualitative approach. The first approach involves the generation of data in quantitative form, which can be subjected to rigorous quantitative analysis. This approach can be further sub-classified into experimental and simulation approaches to research.

We use experimental approach for our case study, and in this case our research is based on quantitative analytical research. We take facts and information already generated by consumers in form of comments and reviews and analyse that using analytical tools to evaluate the data in a more critical fashion. According to Kothari

(2004), in analytical research, the researcher has to use facts or information already available, and analyze these to make a critical evaluation of the material.

4.2 Case Company Description

According to Stake (1995), a case study is the study of particular and complex phenomena of a single case, in order to understand its activity within the circumstances. Kaplan & Duchon (1988), explain that information systems research generally is characterized by a methodology of formulating hypotheses that are tested through controlled experiment or statistical analysis.

We choose to work with Jabra customer reviews intrinsically, to understand the situation and knowledge that implies to it, and emphasize on the case itself. However, the insight can be used to develop understanding of similar field of research that implies social generated content. It is important to learn about the case company, in order to grasp the area of research related to Jabra consumer reviews.

Jabra was founded January 3, 1993, and it is a leading provider of Intelligent Audio Solution, which varying from wired to wireless headsets for mobile users, contact centers and office-based users. In September 2000, Jabra Corporation was acquired by GN Netcom, a division of the Danish company GN Great Nordic. In 2006, GN Netcom consolidated its Contact Center and Office (CC&O) headset division under the Jabra brand. Nowadays, Jabra provides to the market with a large range of headsets and speakerphones being compatible with a broad range of desk phones for offices and call centers.

Understanding the importance of receiving feedback from their consumers, Jabra has opened a platform on their official website, allowing their consumers to rate and express their experience with Jabra products. Consumers are encouraged to provide feedback regarding Jabra products so as to help Jabra improve their products in the future. Nonetheless, consumers are not merely using this platform, but they tend to give comments on other social platforms, for instance, e-commerce websites like Amazon.com or Bestbuy. It, thus, becomes a demand for Jabra analysts to extract helpful information from these communities so as to analyze their consumer opinions and assess their product quality based on consumer experience.

4.3 Dataset Description

The dataset used for the research is the list of consumer reviews about Jabra products on social platforms. The dataset is collected from social platforms (like Amazon and Bestbuy) by Jabra, and an employee from Jabra manually classified these reviews into categories. The dataset provided is in xml format. The total number of records in the dataset is 6,368 records. There are 7 fields in the dataset, including: *store, product_id, review_title, comment, rating, feedback_type* and *key_point.*

Store field has two primary values, which are Amz US and Bestbuy. They are the two primary social platforms where consumers use to express their opinions, so Jabra has made the decision to retrieve data from these platforms.

Product_id is the id of product which consumers are writing reviews about. The value of *product_id* field is an integer. It is not a distinct value in the dataset since one product can receive numerous comments from consumers.

Review_title is the brief title about the consumer reviews. From the Amazon platform, there is no limit on the number of characters that a user can write. However, in the provided dataset, they are text values and have limited size. The maximum number of characters allowed for review_title is 50 characters.

Comment is the major field in the dataset, describing the consumers' opinions about Jabra products. It is a free text value allowing consumers to express their experience without any restriction.

Rating is the rate that the reviewer gives to Jabra's products. The *rating* has values ranging from 1 to 5, when 5 shows that the reviewer is most satisfied with the product, while 1 represents the consumer's dissatisfaction with the product. In Amazon.com platform, the rating stars have following meanings:

- 1 star: *I hate it*
- 2 stars: *I don't like it*
- 3 stars: *It's okay*
- 4 stars: *I like it*
- 5 stars: *I love it*

These two fields *feedback_type* and *key_point* are manually classified by Jabra employee. The *feedback_type* has two values: 1 (positive) and 0 (negative), representing the customer's feeling about the Jabra products. Since there are duplicates of reviews in the dataset with different feedback_type values, we assume that this classification is also for Jabra's internal use.

Lastly, the *key_point* value is how Jabra is classifying comments into sub categories. The sub categories including of the following values: *Comfortability, Connectivity, Functionality, Physical design, Power/Battery, Product durability, Software/App, Sound Quality Rx, Sound Quality Tx* and *Others.*

4.4 Data Analysis Process

In this research, a practical problem from a case company is investigated, analyzed and formulated in a precise manner. The problem under consideration is also motivated to be of general interest, which means it is significant for a wide community of researchers and practitioners. The expected result of this activity is the descriptive knowledge about the properties and scenario of the problem. This section aims to introduce readers with all procedures and processes of data analysis which will be applied in order to answer to the research question.

4.4.1 Data Analysis Process Diagram

The following diagram visualizes entire steps in the experiments. There are three (03) primary procedures in this research, which are: (i) Data preprocessing, (ii) Data analytics experiment with Sentiment Analysis and (iii) Data analytics experiment with Keywords identification. Additionally, the comparison between the suggested approach with the existing approach by Jabra is conducted where applicable in order to evaluate the accuracy and effectiveness of the suggested approach.

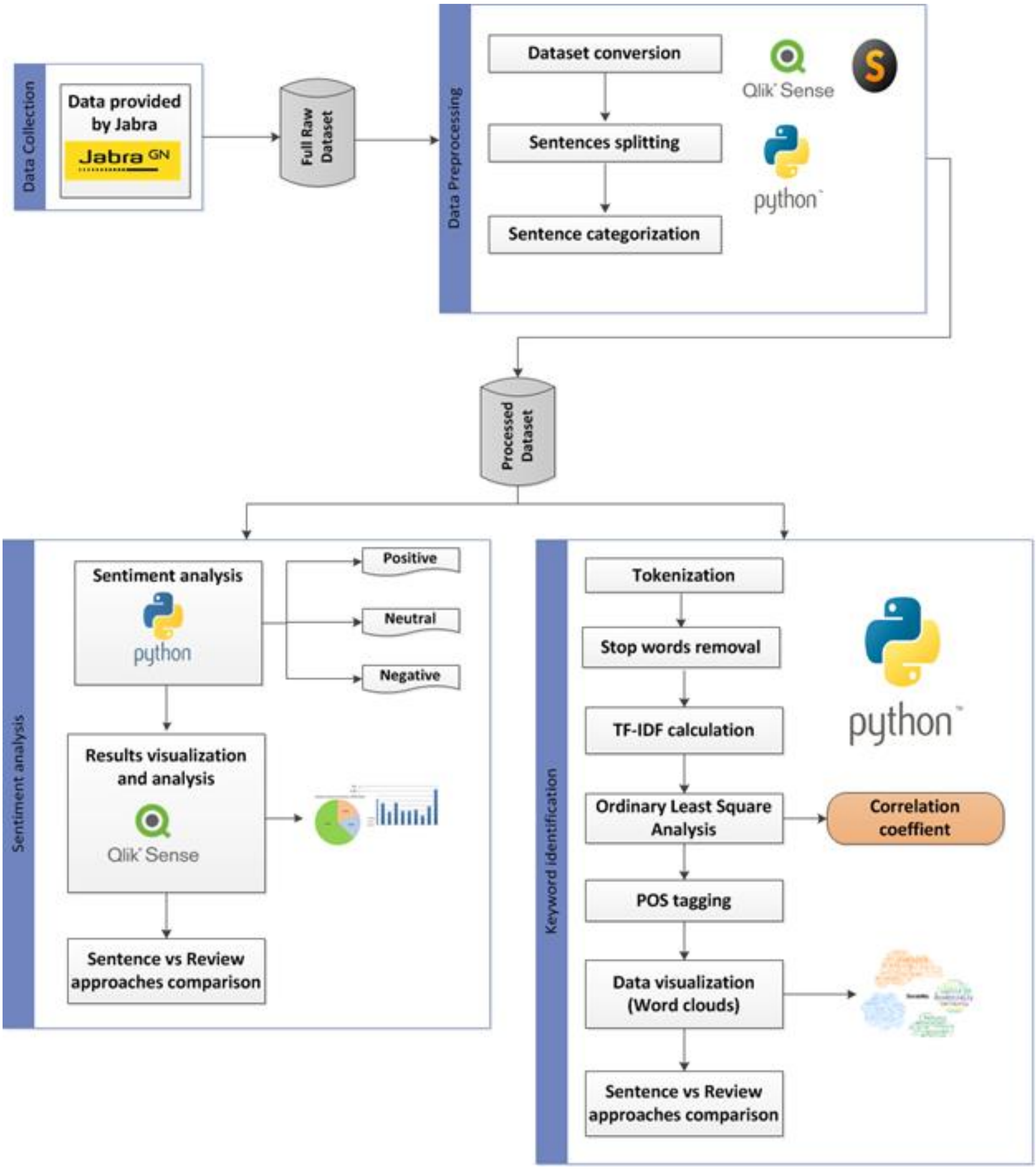


Figure 2 - Data analysis diagram

Process 1: Data preprocessing

Dataset used in this research is collected from two social platforms, Amazon.com and Bestbuy. It is collected and provided by Jabra personnel. The raw data is in .xml format, which is inconvenient for users to perform initial verification, as well as for further analysis in later stage. Thus, we have made the decision that it should be converted into

.csv format. The tool which was used for format conversion is QlikSense Desktop. We also use Sublime to encode the dataset into UTF-8 encode for further analysis.

Secondly, since we aim to suggest an automatic approach to process consumer's reviews instead of the existing manual classification, we decide to perform analysis on sentence level. Thus, reviews must first be split into separate sentences. A small Python program is coded to split reviews into sentences.

The final step of data preprocessing is sentence categorization. Once reviews are split into sentences, only those relevant to Jabra products' features are remained from the sets of sentences. This filtration is based on list of descriptive keywords provided by Jabra. We develop a Python module in order to perform this task.

Procedure 2: Experiment with Sentiment Analysis

Originally, each review from consumers is read and assessed by Jabra personnel. Thus, the review is labelled as either "Positive" (feedback_type = 1) or Negative (feedback_type = 0) on review level and by professional judgement. Nonetheless, this research aims to assess the sentiment value in each sentence of review in an automatic way. Therefore, we conduct sentiment analysis to label these sentences, using Textblob in Python. The result of sentiment analysis provides analysts with Sentiment label, Polarity and Subjectivity values. These values are further applied into QlikSense application to assess Jabra product's quality.

Procedure 3: Experiment with Keywords identification

In OLS regression analysis, we focus on the finding the correlation between words and features. Therefore, the focal point in this experiment is words, or tokens.

Sentences which were remained after filtration step in data preprocessing will be further processed, including tokenization and stop words removal. Also, the TF-IDF value of terms will be calculated, highlighting the importance of words/phrases used by consumers when they provide the feedback about Jabra products. The measures are used for OLS regression analysis to calculate the correlation of the words to the features, so that analysts can identify the highly correlated words with the features. The features are extracted from list of keywords provided by Jabra, which is slightly different from the key_point value in the original dataset.

Afterwards, we implement POS tagging techniques to label descriptive words with their corresponding POS tag. We finally use word clouds as the method to visualize the result and discuss about the result of OLS regression to answer our research and sub research questions.

4.5 Data Collection: Methods and Tools

The dataset is collected and provided by Jabra; therefore, no further works was required from our side in order to retrieve data.

4.6 Data Pre-Processing: Methods, Tools and Techniques

Data pre-processing is the critical part of the preparing data for the experiments. As we propose the sentence approach to analyze consumer opinions, the initial dataset of reviews has to be converted into sentence-based dataset. The following diagram demonstrates the key steps in data pre-processing phase.

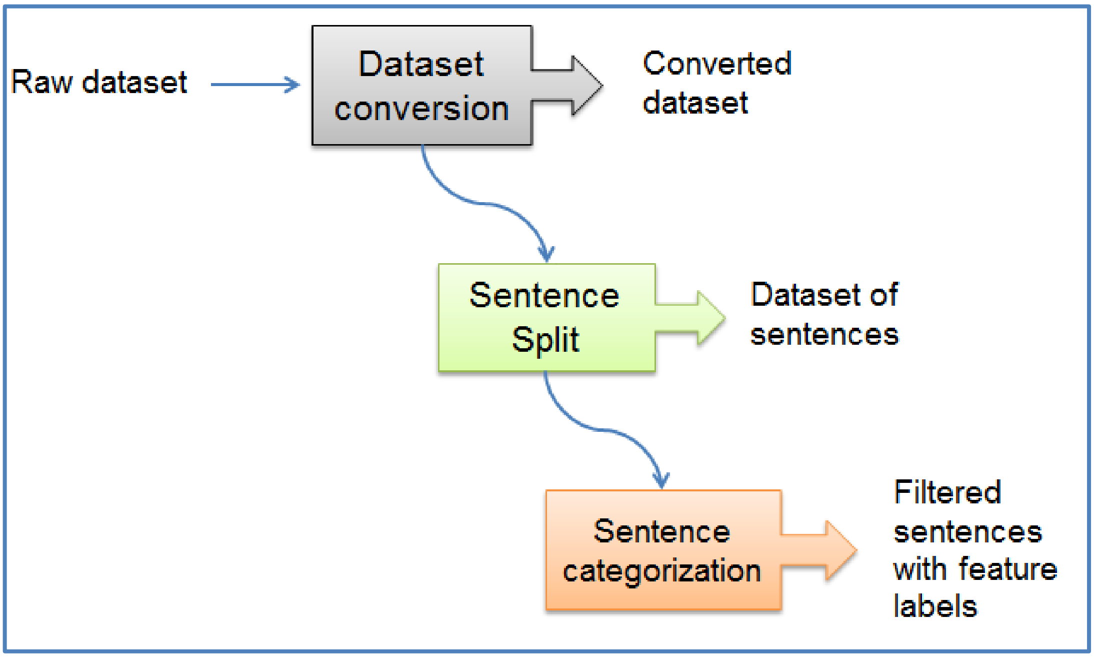


Figure 3 - Data preprocessing diagram

4.6.1 Dataset conversion

The original dataset provided by Jabra is in .xml format. The XML stands for Extensible Markup Language, which is a markup language used to annotate text or add additional information. These annotations are not shown to end-users, so they are not coherent for human use, but are needed by the 'machine' to read and subsequently process the text correctly. In order to perform initial verification on data, we need the dataset in

.csv format. Therefore, we use QlikSense to convert this raw dataset into .csv format.

The following figure is the sample of one consumer reviewer extracted from the original dataset provided by Jabra. Each value is wrapped in tags <> and </>. The complete original dataset is attached in Appendix A.


```
Query1.xml
<?xml version="1.0" encoding="UTF-8"?>
<dataroot xmlns:od="urn:schemas-microsoft-com:officedata" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="Query1.xsd" generated="2017-03-29T15:40:03">
<Query1>
<store>Amz US</store>
<product_id>1</product_id>
<review_title>I regret buying this Jabra Steel (even with huge d</review_title>
<comment>I purchased this Jabra Steel after I ruined my Motorola Whisper trying to quickly rinse off some
muddy water fingerprints. The tiny bit of water ruined the microphone. I would buy the Motorola Whisper again
in a heartbeat if it was water-resistant and currently in production. I would not recommend the Jabra Steel
except for supposedly being water-resistant.

Pros:
Supposedly water and dust resistant. (have not even tried to quickly rinse it under a faucet.)
Works in normal relatively quiet environments.

Cons:
*I would give 2 Stars for volume and noise cancellation, but gave it 3 stars because it works for if its quiet
and may be water-resistant.
*Jabra ear pieces really hurt my ears if I wear them for long periods, even causing sores/scabs in my ears. I
switch ears daily and now have sores in both ears.
*NO volume control, too quiet even with phone turned up to max. Motorola was too loud at max.
*People say I cut out a lot, or they just cant hear me unless I nearly shout.
*Hard to use voice commands to &quot;answer&quot; and actually most often ends up rejecting calls with &quot;
say answer or ignore&quot;.
*The Jabra Assist app makes my phone freeze up for 1-2minutes, mostly when the Jabra Steel is turned on after
the phones bluetooth.
*If it is windy people cannot hear or understand me (havent tried the goofy looking foam cover)(I did not have
an issue with the Motorola)
*USB port cover is pretty tight to try to move out of the way of the charging cord. Im afraid it will break
soon.
* I had never heard of Jabra before my dad gave me a old one which did not work worth a crap, likely why he
did not want it. I purchased this Steel because it was &quot;ruggedized and water-resistant&quot; since I work
outside.</comment>
<rating>3</rating>
<feedback_type>0</feedback_type>
<key_point>Sound Quality Rx</key_point>
</Query1>
```


Figure 4 - Sample of original dataset

For the purpose of converting the dataset into .csv format, a short QlikSense code is deployed.

```
XMLtable:
LOAD
    "store",
    product_id,
    review_title,
    "comment",
    rating,
    feedback_type,
    key_point,
    %Key_dataroot_72A09EACE72FB6DB
FROM [lib://XML file/Query1.xml]
(XmlSimple, table is [dataroot/Query1]);

Store XMLtable into [lib://XML file/Query1.csv] (txt);
```

Figure 5 - QlikSense code for csv conversion

After running the conversion, the dataset is transformed into .csv format, which is easier for human use. By using .csv format, analysts can perform initial verification over the dataset. Some of verification steps include checking for:

- Completeness of data: any missing values which are critical to assessment
- Validation of data: invalid data which cause errors in later stage

4.6.2 UTF-8 encode

UTF-8 encode is a method for character encoding using 8-bit consequences, which is defined by Unicode and originally designed by Ken Thompson and Rob Pike (Yergeau, 1996). UTF stands for Unicode Transformation Format, and the '8' means it uses 8-bit blocks to represent a character.

The main reason that drives us to convert the dataset into UTF-8 encode is due to the fact that UTF-8 is very efficient at encoding plain English text because it only takes up one byte per character, and our dataset is written English text. To convert the dataset into UTF-8, we have used Sublime as a tool to save the file into this encode.

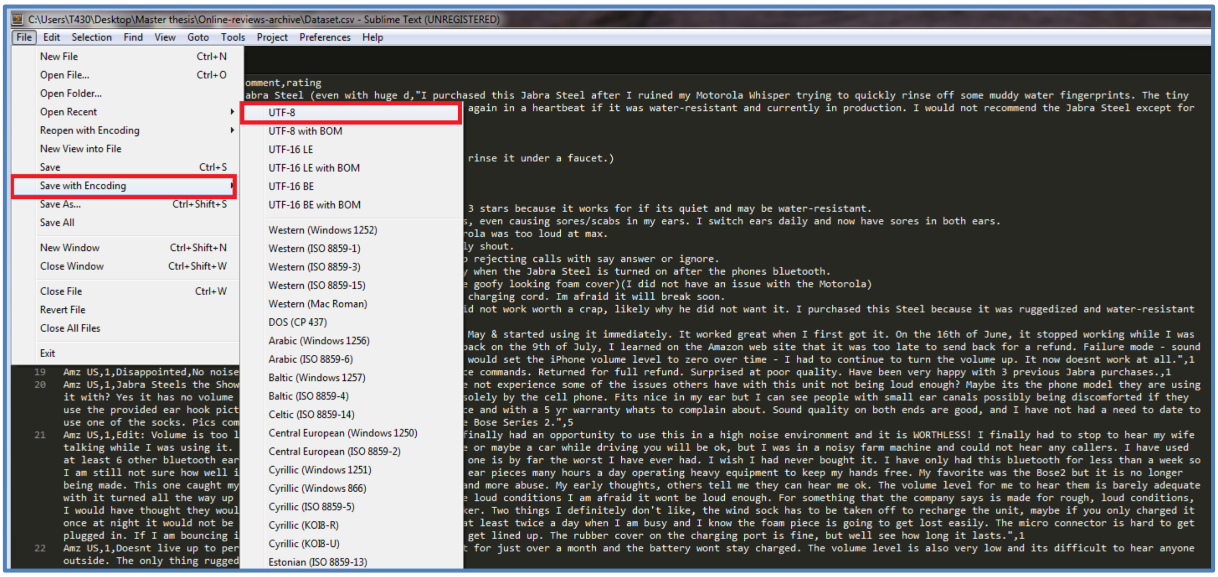


Figure 6 - UTF-8 conversion by Sublime

4.6.3 Reviews into Sentence split

Originally, the dataset consists of reviews of consumers about Jabra products, which usually presented as a short paragraph. These reviews are read and classified by Jabra personnel, so they are based on professional judgement to be labeled with sentiment values, as well as with key_point categories. Nonetheless, we have performed initial assessment over the dataset and notice that consumers tend to give opinions about Jabra products on not merely one feature of the products, but also several features. It has some drawbacks in matching the consumer's rating with specific features, since a weakness in one feature could cause an overall negative impact on ratings. Consequently, it may result in incorrect implication for specific feature quality assessment.

Therefore, the suggested approach for this concern is to analyze consumer reviews on sentence level. It helps analysts find out all possible implying opinions regarding certain features, so as to provide more accurate assessment over product quality. Thus, the purpose of this pre-processing step is creating a new dataset, treating each sentence as a document. The original dataset is transformed into new corpus, which each document is presently a sentence together with its existing fields inherited from its review (store, product_id, review_title and rating). The feedback_type and key_point are, however, not exactly referring to the newly created sentences, so they are eliminated in the new dataset.

The following diagram illustrated how we build up the new dataset based on the original data.

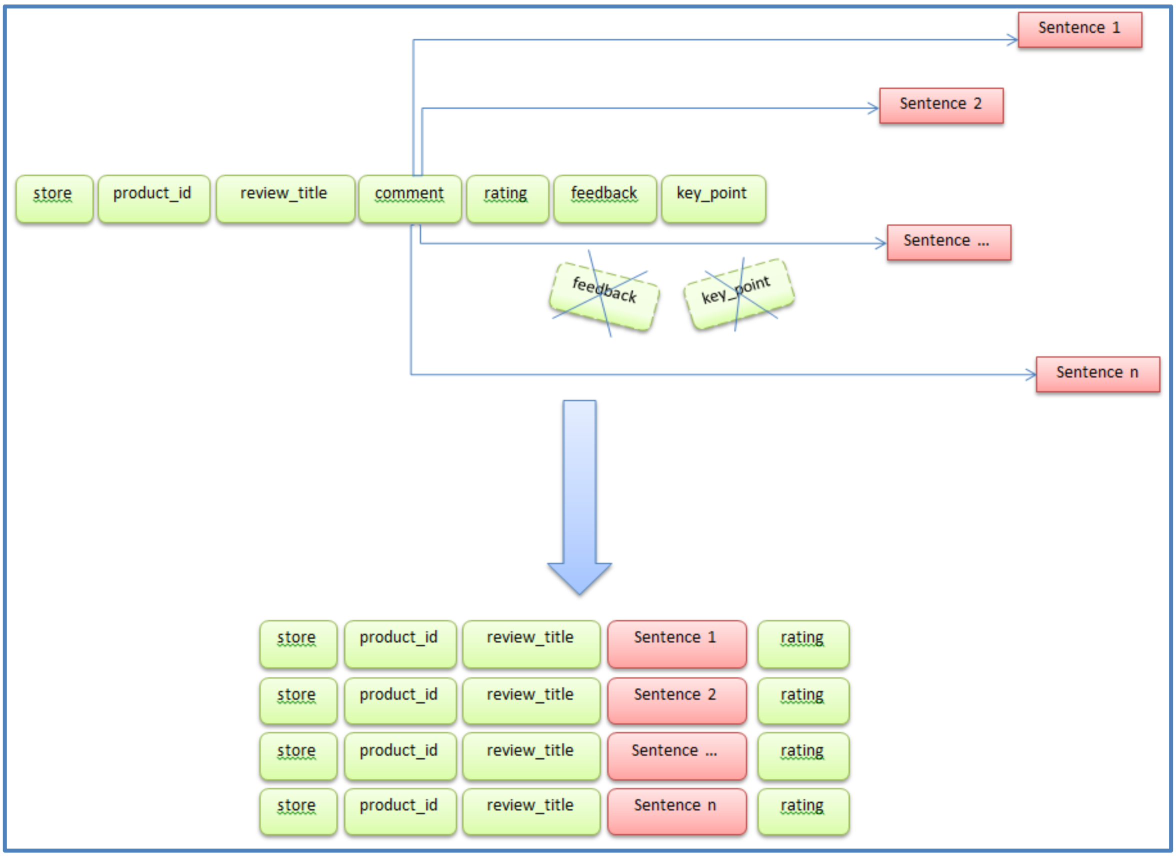


Figure 7 - Sentence split diagram

4.6.3.1 Tools

Python and its subordinate libraries are used for text pre-processing to split a paragraph into sentences. Python is a high-level programming language, which provides constructs intended to enable writing clear programs, so it is straightforward and easy to use. There are also multiple libraries for Python, which can easily be leveraged for text mining purpose.

With the choice of Python as the main scripting language in this dissertation, necessary text mining techniques are provided by the Natural Language Toolkit (NLTK), which is a leading open source platform supporting the analysis of human language data (Bird, Klein and Loper, 2009). NLTK is compatible with both Python 2 and Python 3 versions.

Figure 8 is the screenshot of Python codes for sentence split.

```
import csv
import nltk.data

Sentence_tokenizer = nltk.data.load('tokenizers/punkt/english.pickle')
sourceFile = ("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/Dataset.csv")
targetFile = ("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/SentenceSplit.csv")
with open(sourceFile) as csvfile:
    DataReader = csv.reader(csvfile, delimiter=',',)
    header = next(DataReader)
    with open(targetFile, 'wb') as writeFile:
        TextWriter = csv.writer(writeFile, delimiter=',',quoting=csv.QUOTE_MINIMAL)
        TextWriter.writerow(header)
        for row in DataReader:
            sentences = Sentence_tokenizer.tokenize(row[3])
            beFore = [row[0],row[1],row[2]]
            afTer = [row[4]]
            for i in range (0,len(sentences)):
                TextWriter.writerow(beFore + [sentences[i]] + afTer)
```

Figure 8 - Screenshot of Python code for sentence split

In the code, NLTK is implemented for sentence splitting purpose. NLTK provides the class named as PunktSentenceTokenizer, which uses an unsupervised algorithm to build a model for abbreviation words, collocations, and words that start sentences (Perkins, 2014). Nonetheless, this class must be trained on a large collection of plain text in the target language before it can be used, and the NLTK package data has included a pre-trained Punkt tokenizer for English, so users can implement it in Python codes.

4.6.4 Sentence categorization

As this research aims to explore consumer opinions regarding their experience with products, as well as performing a product quality assessment based on text values from consumer feedback, it is critical to be able to match consumers' experience with specific features of products. Typically, this process is handled by Jabra personnel, thus the feature matching depends on their professional judgement. In this research, an automatic approach is suggested so as to classify reviews into categories.

Because the process of feature extraction from text streams requires certain understanding about Jabra products and their competitive industry, it should include professional expertise in this process. Consequently, we are provided with a list of distinguishing words for categories of features from Jabra (table 1). For further reference, the complete list of keywords is attached in Appendix B.

Sound	Power	Connectivity	Functionality	Design	Durability	Comfortability	Software	Other	Price	Perception	Accessories
Music	Power	BT	Usability	large	waterproof	Fit	application	Heart rate	Cheap	happy	Charger
speech	Battery	Bluetooth	buttons	small	warranty	comfort	APP	Running	expensive	satisfied	cable
Audio	Battery Life	connection	True wireless	shape	sweat proof	uncomfortable	assist	Cycling	overpriced	not happy	ear gels
mono	charging	pairing	Wireless	material	guarantee	poor fit	sport	fitness	competitor	angry	user manual
stereo	capacity	pair	control box	sales box	corrosion	Wings	upgrade	side tone	comparison	unsatisfied	ear buds
microphone	turn on	multiconnection	controller	sales package	overheating	tips	update	Jabra (Brand feedback)	Cheap	recommend	cussions
speaker	power on	multiuse	FM transmission	soft	thermal runaway	Secure	Direct		expensive	return	usb cable
bass	discharge	range	guidance	hard	heating	foam	PC suite	Heart rate	overpriced	unhappy	wings
treble	fast charging	compatibility	guide	flexibility	burning	gel	Firmware	Running		sucks	foam
dropout	battery leakage	Android	quick start guide	accessories	smoke	earbuds	FW	Cycling		disappointed	gels
noise	Power	iOS	out of box	headset	hot	headband				promisse	missing (accessories)
background	Battery	Mac	faulty	glossy	swelling	cable				value	cradle
acoustic shock	Battery Life	PC	quality	robust	peel off	wire					charging case
rattling noise	charging	green	ANC	brittle	strain relief	tight					case
scratch	capacity	purple	noise cancellation	open	break	loose					pouch
level	turn on	blue	mic	closed	crack	neck					wall charger
sound level	power on	red	microphone	speaker housing	damage	pain					Charger

Table 1 - List of keywords by Jabra

The table is first transposed in Microsoft Excel, so that it can revert columns and rows. This step is required to import this table into Python for category classification.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Sound	Music	speech	Audio	mono	stereo	microphone	speaker	bass	treble	dropout	noise
2	Power	Power	Battery	Battery Life	charging	capacity	turn on	power on	discharge	fast charging	battery leakage	
3	Connectivity	BT	Bluetooth	connection	pairing	pair	multiconnection	multiuse	range	compatibility	Android	iOS
4	Functionality	Usability	buttons	True wireless	Wireless	control box	controller	FM transmission	guidance	guide	quick start guide	out of box
5	Design	large	small	shape	material	sales box	sales package	soft	hard	flexibility	accessories	headset
6	Durability	waterproof	warranty	sweat proof	guarantee	corrosion	overheating	thermal runaway	heating	burning	smoke	hot
7	Comfortability	Fit	comfort	uncomfortable	poor fit	Wings	tips	Secure	foam	gel	earbuds	headband
8	Software	application	APP	assist	sport	upgrade	update	Direct	PC suite	Firmware	FW	
9	Other	Heart rate	Running	Cycling	fitness	side tone	Jabra (Brand feedback)					
10	Price	Cheap	expensive	overpriced	competitor	comparison	Cheap	expensive	overpriced			
11	Perception	happy	satisfied	not happy	angry	unsatisfied	recommend	return	unhappy	sucks	disappointed	promisse
12	Accessories	Charger	cable	ear gels	user manual	ear buds	cussions	usb cable	wings	foam	gels	missing (accessories)

Table 2 - List of transposed keywords

Figure 9 displays the Python program to match dataset with list of predefined words by Jabra. The list of keywords is read, and used as the primary criteria for feature matching. If a sentence contains any words in the list of predefined descriptive words, the sentence will be labelled with the categories which the descriptive words belong to. If this sentence contains descriptive words belong to more than one category, it will also be labelled with all possibly related category.

```
sourceFile = ("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/SentenceSplit.csv")
targetFile = ("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/FeatureMap.csv")
featFile = ("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/KeyWords.csv")
import csv
list_of_Keywords = {}
with open(featFile) as csvfile:
    KeyRead = csv.reader(csvfile, delimiter=',',)
    for line in KeyRead:
        line = [x.lower() for x in line]
        list_of_Keywords[line[0]] = []
        for i in range (1,len(line)):
            list_of_Keywords[line[0]].append(line[i])
csvfile.close()
with open(sourceFile) as csvfile:
    DataReader = csv.reader(csvfile, delimiter=',',)
    header = next(DataReader)
    with open(targetFile, 'wb') as writeFile:
        TextWriter = csv.writer(writeFile, delimiter=',',quoting=csv.QUOTE_MINIMAL)
        TextWriter.writerow(header)
        for row in DataReader:
            for k in list_of_Keywords:
                if any (word.lower() in row[3] for word in list_of_Keywords[k]):
                    row.append(k)
            TextWriter.writerow(row)
    writeFile.close()
sourceFile.close()
```

Figure 9 - Python code for Sentence categorization

4.7 Data Analytics: Modeling, Methods and Tools

4.7.1 Experiment I: Sentiment analysis

Initially, the dataset already consists of the sentiment values of comments (field feedback_type), which is classified as 1 (Positive) or 0 (Negative) depending on personal judgement of Jabra personnel. This approach requires human involvement, which has some disadvantages of putting pressure on people to be correct in details, as well as costing time and effort from Jabra personnel.

For the first experiment, we would like to identify the overall evaluation of the public about their experience with Jabra products, using sentiment result from an automatic approach. Therefore, after processing the data, we run the sentiment analysis of the dataset. The goal of this experiment aims to conduct sentiment analysis using tools so as to automate the process and assess the accuracy of the tools. Since the central focus of this paper is not data mining, we do not establish a training dataset, build up classification models and assess the performance of sentiment analysis algorithm. Instead, we have deployed Textblob into our Python program to run the sentiment analysis. Textblob is a Python natural language processing toolkit reusing NLTK corpora which is widely used for text mining, text analytics and text processing.

Figure 10 displays the Python code implemented to retrieved sentiment values for each row in the dataset.

```python
sourceFile = ("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/FeatureMap-formatted.csv")
targetFile = ("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/Feature_SentimentResult.csv")
import csv
from textblob import TextBlob

with open(sourceFile) as csvfile:
    fbDataReader = csv.reader(csvfile, delimiter=',',)
    header = next(fbDataReader)
    header = ['Sentiment Label','Polarity','Subjectivity'] + header
    with open(targetFile, 'wb') as writeFile:
        fbTextWriter = csv.writer(writeFile, delimiter=',',quoting=csv.QUOTE_MINIMAL)
        fbTextWriter.writerow(header)
        for row in fbDataReader:
            tblob = TextBlob(str(row[3]).decode('raw_unicode_escape'))
            if tblob.sentiment.polarity > 0:
                label = 'positive'
            elif tblob.sentiment.polarity < 0:
                label = 'negative'
            else:
                label = 'neutral'
            sentilist = [label,format(tblob.sentiment.polarity, '.2f'),format(tblob.sentiment.subjectivity, '.2f')]
            fbTextWriter.writerow(sentilist + row)
    writeFile.close()
```

Figure 10 - Python code for Sentiment analysis

After running the sentiment analysis on Python, we get sentiment label, “Positive”, “Negative” or “Neutral”; value of polarity and value of subjectivity. If the value of polarity is greater than 0, the status is labelled as "Positive". If the value of polarity is less than 0, the status is labelled as "Negative". Otherwise, this status does not carry any sentiment, or "Neutral". The subjectivity value, on the other hand, helps analysts know whether it is subjective or objective. The value from 0 to 1 measuring the subjectiveness of the text where 0 is objective and 1 is subjective.

The sentiment result provides analysts with various potential applications. First and foremost, it provides an overall assessment about consumer's satisfaction with Jabra products. It is a fast and convenient method for management to understand their consumer contentment. Secondly, dataset with sentiment values can be combined with a Business Intelligence tool like QlikSense to perform detailed analysis. An application built in QlikSense can provide analysts with an ability to select various options, enabling them to assess the product quality in scale of features or products. It helps management to quickly detect weak points in their products so as to plan for future product improvement.

4.7.2 Experiment II: Keywords identification using OLS regression analysis

4.7.2.1 Data stored in database

Dataset after being preprocessed and labelled with categories is loaded into SQLite database. This step is primarily prepared for the second experiment with OLS regression analysis to provide a fast access into database using SQL query. SQLite is a relational database management system, and Python provides binding with the SQLite database.

To interact with database files compatible to SQLite, the DB Browser for SQLite is selected. It is a high quality, visual, open source tool for users and developers to create databases, search, and edit data. It is more user-friendly, which uses a spreadsheet-like interface, and users do not require learning complicated SQL queries to interact with.

4.7.2.2 Tokenization

Tokens, in the context of text analysis, are symbols, words, phrases, or other meaningful elements (Provost & Fawcett, 2013). Accordingly, tokenization is the process of splitting a text stream into tokens based on some certain delimiters. There are several delimiters which can be used as word tokenizers, such as white spaces, comma, punctuation, semicolon, etc. Tokenization works by dividing the text into tokens by separating it on these delimiters. In languages using inter-words space, such as most of Latin languages, tokenization approach is straightforward. However, in other languages where there is an ambiguous boundary among words, such as Chinese, Korean or Arabic languages, it is more complicated to do tokenization.

Since the language used in the dataset is English, Python and its subordinate libraries are used in the research. NLTK provides users with a class named word_tokenize to do the tokenization. It helps tokenize a string to split off punctuation other than periods. In the Python code, the word tokenizer is declared as a function because it will be applied into TfidfVectorizer.

```python
import sqlite3
import sys
import numpy as np
import nltk
from sklearn.externals import joblib

connection = sqlite3.connect("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/MasterThesis.db")
connection.text_factory = str
cur = connection.cursor()

def tokenize(text):
    tokens = nltk.word_tokenize(text)
    return tokens

cur.execute('SELECT comment FROM JabraReviews')
returned_reviews = cur.fetchall()
lineCount = len(returned_reviews)
print "Size of corpus: " + str(lineCount)
returned_reviews_str = []
returned_reviews_str =["".join(review) for review in returned_reviews]

# Calculate tf-idf for the whole corpus, 1 word with stop word removal
from sklearn.feature_extraction.text import TfidfVectorizer
tf = TfidfVectorizer(analyzer='word',tokenizer = tokenize, min_df = 1, stop_words ="english")
tfidf_matrix =  tf.fit_transform(returned_reviews_str) #Learn vocabulary and idf, return term-document matrix.
feature_names = tf.get_feature_names() #Array mapping from feature integer indices to feature name
print ("Number of phrases: " + str(len(feature_names)))

#Retrieve only idf score to save into file
idf=tf._tfidf.idf_
p = zip(tf.get_feature_names(), idf)
p.sort(key = lambda t: t[1], reverse = True)

joblib.dump(p, "idf.pkl")
```

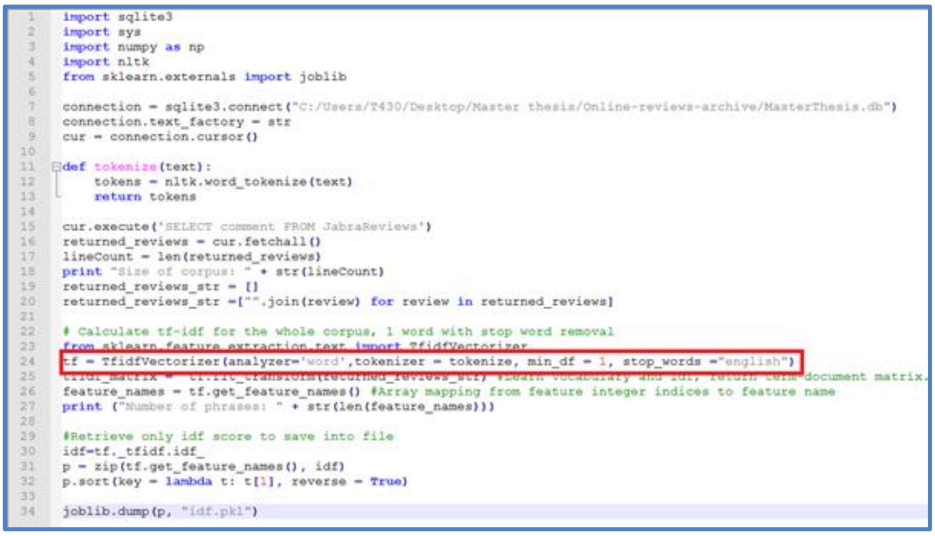
Figure 11 - Python code for Tokenization

4.7.2.3 Stop word removal

In the area of natural language processing, stop words are words that are filtered out before further text mining techniques are applied. Stop words refer to words which are very common in language, for example, English has several stop words like "*I*", "*the*", "*you*" and so on. From the perspective of data analysis, a term occurring in every document is unimportant; for instance, it would not help distinguish any classes, or serve as a basis for any cluster (Provost & Fawcett, 2013). Consequently, these words should be removed in data preprocessing stage.

Regarding the second experiment with quality assessment using OLS regression analysis, the purpose is to identify the most relevant words that illustrate the consumers' opinions. Therefore, all stop words should be removed to enhance the accuracy of OLS regression. TfidfVectorizer has the parameter allowing users to select to ignore stop words when build tfidf matrix. Thus, we use this stop_word parameter of TfidfVectorizer to exclude all English stop words.

The following screenshot shows how tokenization and stop word removal parameters are applied in TfidfVectorizer in Python.



```
import sqlite3
import sys
import numpy as np
import nltk
from sklearn.externals import joblib

connection = sqlite3.connect("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/MasterThesis.db")
connection.text_factory = str
cur = connection.cursor()

def tokenize(text):
    tokens = nltk.word_tokenize(text)
    return tokens

cur.execute('SELECT comment FROM JabraReviews')
returned_reviews = cur.fetchall()
lineCount = len(returned_reviews)
print "Size of corpus: " + str(lineCount)
returned_reviews_str = []
returned_reviews_str =["".join(review) for review in returned_reviews]

# Calculate tf-idf for the whole corpus, 1 word with stop word removal
from sklearn.feature_extraction.text import TfidfVectorizer
tf = TfidfVectorizer(analyzer='word',tokenizer = tokenize, min_df = 1, stop_words ="english")
tfidf_matrix = tf.fit_transform(returned_reviews_str) #Learn vocabulary and idf, return term-document matrix.
feature_names = tf.get_feature_names() #Array mapping from feature integer indices to feature name
print ("Number of phrases: " + str(len(feature_names)))

#Retrieve only idf score to save into file
idf=tf._tfidf.idf_
p = zip(tf.get_feature_names(), idf)
p.sort(key = lambda t: t[1], reverse = True)

joblib.dump(p, "idf.pkl")
```

Figure 12 - Python code for Stop word removal

4.7.2.4 TF-IDF calculation

TF-IDF is a common way to judge the topic of an article by the words it contains. With TF-IDF, words are given weight – TF-IDF measures relevance, not frequency. That is, word counts are replaced with TF-IDF scores across the whole dataset.

In order to get the TF-IDF values, we have deployed sklearn, which is a free software machine learning library for the Python programming language. In sklearn, there is a pre-built feature selection module, named TfidfVectorizer, using for converting a collection of raw documents to a matrix of TF-IDF features. We use this module to retrieve the TF-IDF values of all words in the corpus. This module has several parameters, allowing users to adjust the selection accordingly.

We are interested in the following parameters when processing the dataset:

- analyzer: whether analysts want to process the dataset as features of words or character n-grams.
- tokenizer: for tokenizing the sentences into tokens
- ngram_range: whether analysts want to process unigrams or ngrams.
- stop_words: whether analysts want to keep or remove stop words from the dataset.
- lowercase: to convert all characters into lowercase before tokenizing.

The result of TfidfVectorizer is a matrix with all TF-IDF values of words in the dataset. The TF-IDF values are then extracted for further analysis. The following figure displays the Python code implementing sklearn and calculating the TF-IDF values.

```
import sqlite3
import sys
import numpy as np
from sklearn.externals import joblib

connection = sqlite3.connect("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/MasterThesis.db")
connection.text_factory = str
cur = connection.cursor()

cur.execute('SELECT comment FROM JabraReviews')
returned_reviews = cur.fetchall()
lineCount = len(returned_reviews)
print "Size of corpus: " + str(lineCount)
returned_reviews_str = []
returned_reviews_str =["".join(review) for review in returned_reviews]

# Calculate tf-idf for the whole corpus, 1 word with stop word removal
from sklearn.feature_extraction.text import TfidfVectorizer
tf = TfidfVectorizer(analyzer='word', min_df = 1, stop_words ="english")
tfidf_matrix =  tf.fit_transform(returned_reviews_str) #Learn vocabulary and idf, return term-document matrix.
feature_names = tf.get_feature_names() #Array mapping from feature integer indices to feature name
print ("Number of phrases: " + str(len(feature_names)))

#Retrieve only idf score to save into file
idf=tf._tfidf.idf_
p = zip(tf.get_feature_names(), idf)
p.sort(key = lambda t: t[1], reverse = True)

joblib.dump(p, "idf.pkl")
```

Figure 13 - Python code for TF-IDF calculcation

4.7.2.5 OLS regression analysis

The choice of OLS regression analysis is due to the following reasons: (i) it is one of the most basic and most commonly used prediction techniques known to humankind, with applications in fields as diverse as statistics, finance, medicine, economics, and psychology. (Dismuke & Lindrooth, 2006), (ii) It is easy to implement on a computer using commonly available algorithms from linear algebra. (iii) It requires minimal understanding at basic level for users to apply and interpret results mathematically.

In an OLS regression analysis, it is critical to firstly identify the dependent variables (response variables) and independent variables (explanatory variables). The dependent variables are variables which represent the output or outcome whose variation is being studied. The independent variables, on the other hand, are variables which represent inputs or causes, or potential reasons for variation. Usually, they can be implied as the causality: Independent variables (X) -> Dependent variables (Y)

For the experiment, the aim is to identify the correlation between words (weighted by TF-IDF values) and product features (represented by key_point value). Thus, it results in the decision that the independent variables are the tfidf values, which is the result extracted from TfidfVectorizer. The TF-IDF matrix contains of words and their corresponding TF-IDF values in the corpus. To simplify the experiment, the dependent variables are built based on a binary mode: if the sentence mentions about one specific feature under experiment, the dependent value is 1, and if not, the value is 0. By doing so, a set of binary values for dependent variables is built and the OLS regression analysis can be conducted on these variables. The result of OLS regression analysis will display how words are correlated with the defined feature.

The experiment can be extended to the context of rating. Instead of just setting the value of dependent variable as 0 or 1 depending on the feature mentioned in the sentence, we can also combine with the rate of review. It will help analyze negative reviews about products, and based on the results, analysts can identify words/expressions that highly related to negative reviews specifically for one feature.

To run OLS regression analysis in Python, we need to import statsmodels, which is a Python module supporting users to explore data, estimate statistical models, and perform statistical tests. The following figure is a screenshot of Python program for OLS regression analysis:

```
import operator
import json
from collections import Counter
import myTokenizer #myTokenizer.py
from nltk.corpus import stopwords
import string
from nltk import bigrams
from nltk.util import ngrams
import sqlite3
import sys
import csv
import numpy as np
import statsmodels.api as smf
from sklearn.externals import joblib

clf = joblib.load('idf.pkl')
tf_idf_dict = dict ()
for line in clf:
    # temp = line.split(",")
    tf_idf_dict [line[0]] = float(line[1])

connection = sqlite3.connect("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/MasterThesis.db")
connection.text_factory = str
cur = connection.cursor()

cur.execute('SELECT comment,key_points FROM JabraReviews')
returned_reviews = cur.fetchall()
lineCount = len(returned_reviews)
print "Size of corpus: " + str(lineCount)
```

Figure 14 - Python code for OLS regression analysis

To extend the experiment in order to analyze reviews with negative feedback from users (with rating = 1), the code is modified as follows:

```
connection = sqlite3.connect("C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/MasterThesis.db")
connection.text_factory = str
cur = connection.cursor()

cur.execute('SELECT comment,key_points,rating FROM JabraReviews')
returned_reviews = cur.fetchall()
lineCount = len(returned_reviews)
print "Size of corpus: " + str(lineCount)
for key_point in list_of_keypoints:
    print key_point
    dependent_var = []
    for review in returned_reviews:
        if ("".join(review[1]) == key_point) and (int(review[2]) == 1):
            dependent_var.append(1)
        else:
            dependent_var.append(0)
```

Figure 15 - Python code for extended OLS regression

4.7.2.6 POS Tagging

Understanding POS tag of a word provides insights into the context of a free text stream. For the purpose of quality assessment, POS tagging can help highlight the search result by focusing on relevant part of speech. We use POS tagging technique in the experiment in order to classify and group words into their corresponding POS groups, which will be further analyzed. Once words are grouped according to their POS categories, analysts can easily identify and detect any issues with the Jabra products' features.

POS tagging in NLTK is supported by the library nltk.pos_tag. The following screenshot demonstrate how nltk.pos_tag has been deployed in the experiment to tag the tokens.

```
import nltk
import csv
import os

def append_id(filename):
    parts = filename.split('.')
    return "".join(parts[:-1])+ '_POStag' + '.' + parts[-1]

directory = 'C:/Users/T430/Desktop/Master thesis/Online-reviews-archive/OLS result'
for root,dirs,files in os.walk(directory):
    for sourceFile in files:
       if sourceFile.endswith(".csv"):
            with open(sourceFile) as csvfile:
                KeyRead = csv.reader(csvfile, delimiter=',',)
                header = next(KeyRead)
                header.append ("POS tag")
                targetFile = append_id(sourceFile)
                with open(targetFile,'wb') as writeFile:
                    TextWriter = csv.writer(writeFile, delimiter=',',quoting=csv.QUOTE_MINIMAL)
                    TextWriter.writerow(header)
                    for line in KeyRead:
                        tag = nltk.tag.pos_tag([line[0]])
                        line.append(tag[0][1])
                        TextWriter.writerow(line)
                writeFile.close()
```

Figure 16 - Python code for POS Tagging

These POS tagging values of tokens are helpful in result visualization. Based on categories classified by POS tagging, we can develop word clouds to visualize the OLS regression analysis results, grouping into sub-categories of POS tagging values. It provides Jabra managers with an overview of results of implementing text mining techniques into consumer reviews to assess Jabra product quality.

4.7.2.7 Data visualization

We generate the word clouds based on the result from OLS regression analysis in order to visualize the relationship among unit of languages used in the reviews and the defined product attributes. However, unlike the common word clouds which use the word frequency as the main scale for font size, this research uses the coefficient values as the base to generate word clouds. Therefore, words with higher coefficient values will appear bigger in the word clouds.

Additionally, our word clouds use different colors for representing different POS. It provides different dimensional views for audience, especially for Jabra managers who aim to use the results for product quality management. Nouns, adjectives and verbs are grouped and visualized into 3 corresponding clouds with different colors (blue for nouns, green for adjectives and orange for verb). The remaining words belong to one group. So, for each product attribute, there are 4 small clouds representing the correlation strengths.

The word clouds are generated by utilizing the word cloud Python codes from Andreas Mueller from Github[1] . It allows us the ability to customize the word cloud and adjust the codes according to our intention.

4.7.3 Comparison between review approach versus sentence approach

The purpose of this process is to compare the result from sentence approach with ones of review approach. It is essential to perform this comparison, since one of the goals of this research is providing an automatic approach towards the current manual methods deployed by Jabra team. Instead of the traditional approach of manually reading and classifying reviews performed by Jabra personnel, this research recommends the case company with the new method of accessing the review on sentence level, classifying them into their corresponding categories of features, and extracting useful information from data. They are all done automatically by applying text mining techniques. Therefore, the comparison between these two approaches helps analysts to evaluate the accuracy and effectiveness of suggested approach in resolving the business concern.

These approaches (review-level and sentence-level) can be compared where applicable in term of the following aspects:

- How accurate in labelling sentences with sentiment values ("Positive", "Neutral" and "Negative") ?
- How accurate in categorizing sentences into features based on predefined keywords?

Answers to these questions can provide a good measurement for the effectiveness of the suggested approach. Also, the result of comparison can provide the case company with suggestions when applying this approach in future product assessment.

[1] https://github.com/amueller/word_cloud

4.8 Dataset Limitations

As explained above in the data set description, the data consists of 6,368 total number of records. The data as the source indicated mainly consists of product reviews and customer opinions about Jabra electronic headsets and headphones. The dataset, however, consists only of customer product reviews from Amazon.com and Bestbuy social platforms. It has not included reviews from consumers on other social platforms, such as Facebook, Twitter or Jabra official webpage. Therefore, the result of this research may not address concerns of all Jabra consumers.

The second limitation which may have impacts on this research is the key_point categories used in this dataset is slightly different from those in list of keywords for feature matching provided by Jabra. The differences in these two types of classification restrict us from evaluating the accuracy of the suggested approach since they are not exactly matched.

Another limitation is the length of review title which is restricted to maximum 50 characters long. It is, however, not restricted in the original review on e-commerce platforms. If the dataset contains of the complete review title, it can provide an overall quick assessment for product quality.

Last but not the least, compared to an online review, it is noted that the dataset does not include statistics about the number of reviewers who find the review helpful. Comments with more "likes" meaning that it receives more approval from the public, and the more "likes" a comment received the more exposure it gets to other consumers. Therefore, a negative comment which receives a lot of "likes" from the public can turn into top comment and impacts on consumer's choice. It thus requires some algorithm to take it into consideration when addressing consumer opinions. Our research treats all reviews equivalently since there is no statistic value to it, therefore, we are not able to evaluate the impact of popular reviews on the community.

By understanding these aforementioned dataset limitations, we believe that we can assess the results of these experiments more accurately.

4.9 Data Analytics: Alternate Methods & Cost/Benefit Analysis

The choice of Python and their subordinate libraries as the primary tool for data analysis in this research has been carefully considered in the initial stage of the research. There are alternative programming languages for solving the research questions, like Java or R programming. However, considering the nature of this dissertation is text mining, Python is selected as the main scripting language because of the following reasons: (i) There are plenty of ready-made libraries for Python, including Natural Language Toolkits (like NLTK), sklearn and SQLite database interaction, which ease the effort for the implementation phase and (ii) Python syntax is simpler, more intuitive and reusable in larger projects. (iii) Data collected from social platforms is in the form of plain text. Python is the leading language in the domain of text processing and analysis, which is widely used in almost every component of this dissertation.

In addition, for business intelligence, there are several options available for analysts such as Tableau, SAS etc. The reason for choosing QlikSense Desktop for this research is due to the fact that QlikSense provide users with the ability to customize applications by script, enabling more flexibility in developing applications. Also, although alternatives like Tableau or SAS can have software version for students, they have limited time period for free access for users. QlikSense Desktop is free software available for all users, so it is preferable in case of cost/benefit analysis.

CHAPTER 5

Experiments setup, Results and Analysis

5. Experiments setup, Results and Analysis

This chapter aims at demonstrating the experiments performed in the research, including the experiment setup, results and analysis. There are 04 (four) subsections of this chapter which are organized as follows: Section 5.1 is the result from data preprocessing procedure. Section 5.2 is the experiment with Sentiment analysis. Section 5.3 is the experiment with OLS regression analysis. Lastly, section 5.4 is the valuable outcomes from the experiment results.

5.1 Data Preprocessing

This task employs the four-step process discussed in section 4.6 to preprocess the dataset, which transform the review-based dataset into sentence-based dataset for data analytics experiments. The results for each step are discussed below.

5.1.1 Dataset conversion & UTF-8

By applying methods as introduced in section 4.6, we have received the dataset as demonstrated in figure 17. The dataset consists of 6,368 records, with 7 primary data fields.

	A	B	C	D	E	F	G	H	I	J	K
1	store	product_i	review_title	comment	rating	feedback_	key_point	%Key_dataroot_72A09EACE72FB6DB			
2	Amz US	1	I regret buying this Jabra Steel (even with huge d	I purchased this Jabra Steel after I	3	0	Sound Quality Rx	0			
3	Amz US	1	Very poor quality - failed in 30 days	I ordered this on 15 May & started u	1	0	Sound Quality Rx	0			
4	Amz US	1	Disappointed	No noise cancelling; did not recogni	1	0	Sound Quality Rx	0			
5	Amz US	1	Jabra Steels the Show	Bought locally for same price. I have	5	1	Sound Quality Rx	0			
6	Amz US	1	Edit: Volume is too low in noisy environments. DO	Edit: I finally had an opportunity to	1	0	Sound Quality Rx	0			
7	Amz US	1	Doesnt live up to performance claims	I have had this product for just over	1	0	Sound Quality Rx	0			
8	Amz US	1	held up nicely. I called customer care and they ha	I am impressed with the noise redu	2	1	Sound Quality Rx	0			
9	Amz US	1	Simply Awesome	Week One product review, Ive got	5	1	Sound Quality Rx	0			
10	Amz US	1	Simply Awesome	Week One product review, Ive got	5	0	Sound Quality Rx	0			
11	Amz US	1	Clarity for people on the other end is excellent s	For starters the steel is my first	4	1	Sound Quality Rx	0			
12	Amz US	1	The size and the fit is perfect. It does take some	The size and the fit is perfect. It doe	4	0	Sound Quality Rx	0			
13	Amz US	1	Everything except call waiting	The connectivity was easy. Sound q	3	1	Sound Quality Rx	0			
14	Amz US	1	Dont Let 5 Year warranty fool you...	It may have 5 year warranty but I ha	2	0	Sound Quality Rx	0			
15	Amz US	1	Well Constructed, but Lacks the Volume Needed fo	The Jabra Steel Ruggedized	3	0	Sound Quality Rx	0			
16	Amz US	1	One Star	Speaker volume is tooooo loww	1	0	Sound Quality Rx	0			
17	Amz US	1	comfortable design but not my first choice for a h	The Jabra Steel Bluetooth headset.	3	1	Sound Quality Rx	0			
18	Amz US	1	A very light, but rugged headset at last...	As a good rugged bluetooth	4	1	Sound Quality Rx	0			
19	Amz US	1	I have to say Im pretty impressed with this Jabra	I have to say im pretty impressed	5	1	Sound Quality Rx	0			
20	Amz US	1	A Blue-Tooth Ear piece that requires a smart phone	This ear piece uses three	4	1	Sound Quality Rx	0			
21	Amz US	1	Good sound quality - poor button functionality	Prior to trying the Jabra Steel, Ive	3	1	Sound Quality Rx	0			
22	Amz US	1	Needs some improvements	This is initially a very impressive	2	0	Sound Quality Rx	0			
23	Amz US	1	Good Product	Connectivity is really good, ranges a	5	1	Sound Quality Rx	0			
24	Amz US	1	Really close!!	I have used Bluetooth headsets	4	1	Sound Quality Rx	0			
25	Amz US	1	The call quality is great. The music audio quality	The call quality is great. The music a	4	1	Sound Quality Rx	0			
26	Amz US	1	Firmware Update Greatly Improves Volume	Bought this headset for work and	4	0	Sound Quality Rx	0			
27	Amz US	1	Works as Intended	This isnt the most sleek executivey	5	1	Sound Quality Rx	0			
28	Amz US	1	Sound quality good, but volume nearly inaudible re	Sound quality good, but volume	4	1	Sound Quality Rx	0			
29	Amz US	1	Sound quality good, but volume nearly inaudible re	Sound quality good, but volume	4	0	Sound Quality Rx	0			
30	Amz US	1	Some great features but fails as a functional head	Purchased this from Best Buy and	1	0	Sound Quality Rx	0			
31	Amz US	1	This is just about the best	Ill be the first to admit that Im not	5	1	Sound Quality Rx	0			
32	Amz US	1	Dont buy if you like to be able to hear your calls	Tough and easy to use. However, te	1	0	Sound Quality Rx	0			

Figure 17 - Screenshot of table from dataset conversion

After performing the initial dataset validation, we remark that there is no missing or invalid value in the dataset. Therefore, the dataset is applicable for further analysis.

5.1.2 Sentence split

Figure 18 displays the result of sentence splitting step. A review is divided into a set of sentences, and each sentence is a record in the dataset. One record consists of 05 (five) fields: store, product_id, review_title, comment and rating. From the original set of reviews, we have split it into 16,403 sentences.

	A	B	C	D	E	F
1	store	product_id	review_title	comment	rating	
2	Amz US	1	I regret buying this Jabra Steel (even with huge d	I purchased this Jabra Steel after I ruined my Motorola Whisper trying to quic	3	
3	Amz US	1	I regret buying this Jabra Steel (even with huge d	The tiny bit of water ruined the microphone.	3	
4	Amz US	1	I regret buying this Jabra Steel (even with huge d	I would buy the Motorola Whisper again in a heartbeat if it was water-resista	3	
5	Amz US	1	I regret buying this Jabra Steel (even with huge d	I would not recommend the Jabra Steel except for supposedly being water-re	3	
6	Amz US	1	I regret buying this Jabra Steel (even with huge d	Pros:	3	
7	Amz US	1	I regret buying this Jabra Steel (even with huge d	(have not even tried to quickly rinse it under a faucet.)	3	
8	Amz US	1	I regret buying this Jabra Steel (even with huge d	Works in normal relatively quiet environments.	3	
9	Amz US	1	I regret buying this Jabra Steel (even with huge d	Cons:	3	
10	Amz US	1	I regret buying this Jabra Steel (even with huge d	*Jabra ear pieces really hurt my ears if I wear them for long periods, even cau	3	
11	Amz US	1	I regret buying this Jabra Steel (even with huge d	I switch ears daily and now have sores in both ears.	3	
12	Amz US	1	I regret buying this Jabra Steel (even with huge d	*NO volume control, too quiet even with phone turned up to max.	3	
13	Amz US	1	I regret buying this Jabra Steel (even with huge d	Motorola was too loud at max.	3	
14	Amz US	1	I regret buying this Jabra Steel (even with huge d	*People say I cut out a lot, or they just cant hear me unless I nearly shout.	3	
15	Amz US	1	I regret buying this Jabra Steel (even with huge d	*Hard to use voice commands to answer and actually most often ends up reje	3	
16	Amz US	1	I regret buying this Jabra Steel (even with huge d	*The Jabra Assist app makes my phone freeze up for 1-2minutes, mostly whe	3	
17	Amz US	1	I regret buying this Jabra Steel (even with huge d	*If it is windy people cannot hear or understand me (havent tried the goofy	3	
18	Amz US	1	I regret buying this Jabra Steel (even with huge d	Im afraid it will break soon.	3	
19	Amz US	1	I regret buying this Jabra Steel (even with huge d	* I had never heard of Jabra before my dad gave me a old one which did not v	3	
20	Amz US	1	I regret buying this Jabra Steel (even with huge d	I purchased this Steel because it was ruggedized and water-resistant since I w	3	
21	Amz US	1	Very poor quality - failed in 30 days	I ordered this on 15 May & started using it immediately.	1	
22	Amz US	1	Very poor quality - failed in 30 days	It worked great when I first got it.	1	
23	Amz US	1	Very poor quality - failed in 30 days	On the 16th of June, it stopped working while I was waiting for a plane at LAX	1	
24	Amz US	1	Very poor quality - failed in 30 days	When I came back on the 9th of July, I learned on the Amazon web site that it	1	
25	Amz US	1	Very poor quality - failed in 30 days	Failure mode - sound started being quite tinny sounding, with diminishing vc	1	
26	Amz US	1	Very poor quality - failed in 30 days	Then it would set the iPhone volume level to zero over time - I had to contin	1	
27	Amz US	1	Very poor quality - failed in 30 days	It now doesnt work at all.	1	
28	Amz US	1	Disappointed	No noise cancelling; did not recognize any voice commands.	1	
29	Amz US	1	Disappointed	Returned for full refund.	1	
30	Amz US	1	Disappointed	Surprised at poor quality.	1	
31	Amz US	1	Disappointed	Have been very happy with 3 previous Jabra purchases.	1	

Figure 18 - Screenshot of table from sentence split result

5.1.3 Sentence categorization

Figure 19 is the result of feature matching step. Each row in the dataset is labelled with product features which the sentence may mention about.

store	product_i	review_title	comment	rating			
Amz US	1	comfortable design but not my first choice for a h	The App also contains a locator feature to find your headset and digital manuals.	3	design		
Amz US	1	comfortable design but not my first choice for a h	Summary: While this headset is very comfortable to wear and sounds good on my end, it may not be good for those with whom you are talking.	3	sound	design	comfortability
Amz US	1	comfortable design but not my first choice for a h	The short battery life means you will have to charge it mid day.	3	power		
Amz US	1	comfortable design but not my first choice for a h	Having to use space on the phone for a separate App to use the device is also a downer for me.	3			
Amz US	1	comfortable design but not my first choice for a h	This would not be my first choice of a headset.	3	design		
Amz US	1	A very light, but rugged headset at lastâ€¦	As a good rugged bluetooth headset, yes, this is great.	4	connectivity	design	
Amz US	1	A very light, but rugged headset at lastâ€¦	There is excellent noise cancellation/suppression and air/wind noise suppression with this one.	4	sound	functionality	
Amz US	1	A very light, but rugged headset at lastâ€¦	Plus it is not bulky or big for a rugged headset that you might expect.	4	design		
Amz US	1	A very light, but rugged headset at lastâ€¦	This is much smaller than some of the bluetooth dual ear headsets in the market.	4	connectivity	design	

Figure 19 - Screenshot of table from sentence categorization result

Sentences containing no descriptive words will not be labelled. We make an assumption that if a sentence does not contain any descriptive words from the list provided by Jabra, it will not carry any value for mining, so it will be eliminated from the dataset. From the set of 16,403 sentences, there are 6,244 sentences which are not categorized, accounting for 38.07% of total records.

For one sentence belongs to more than one category and has more than one label, it is added into a new record for each category in the table. It results in the final dataset for further analysis as displayed in figure 20. The entire new dataset consists of 17,314 records, and it is attached in Appendix C.

	A	B	C	D	E	F
1	store	product_id	review_title	comment	rating	key_points
2	Amz US	1	I regret buying this Jabra Stee	The tiny bit of water ruined th	3	sound
3	Amz US	1	I regret buying this Jabra Stee	I would not recommend the J	3	perception
4	Amz US	1	I regret buying this Jabra Stee	Cons:	3	sound
5	Amz US	1	I regret buying this Jabra Stee	*NO volume control, too quie	3	sound
6	Amz US	1	I regret buying this Jabra Stee	Motorola was too loud at max	3	sound
7	Amz US	1	I regret buying this Jabra Stee	*The Jabra Assist app makes n	3	sound
8	Amz US	1	I regret buying this Jabra Stee	*If it is windy people cannot	3	power
9	Amz US	1	I regret buying this Jabra Stee	Im afraid it will break soon.	3	durability
10	Amz US	1	Very poor quality - failed in 3(	I ordered this on 15 May & sta	1	connectivity
11	Amz US	1	Very poor quality - failed in 3(	Failure mode - sound started	1	sound
12	Amz US	1	Very poor quality - failed in 3(	Then it would set the iPhone	1	sound
13	Amz US	1	Disappointed	No noise cancelling; did not re	1	sound
14	Amz US	1	Disappointed	Surprised at poor quality.	1	functionality
15	Amz US	1	Disappointed	Have been very happy with 3	1	perception
16	Amz US	1	Jabra Steels the Show	Yes it has no volume controls	5	sound
17	Amz US	1	Jabra Steels the Show	Fits nice in my ear but I can se	5	connectivity
18	Amz US	1	Jabra Steels the Show	Battery life and features are n	5	durability
19	Amz US	1	Jabra Steels the Show	Sound quality on both ends ar	5	functionality
20	Amz US	1	Edit: Volume is too low in noi	Edit: I finally had an opportun	1	sound
21	Amz US	1	Edit: Volume is too low in noi	If you are going to use it in an	1	connectivity
22	Amz US	1	Edit: Volume is too low in noi	I have used at least 6 other bl	1	connectivity
23	Amz US	1	Edit: Volume is too low in noi	I have only had this bluetooth	1	connectivity
24	Amz US	1	Edit: Volume is too low in noi	I use bluetooth ear pieces ma	1	connectivity
25	Amz US	1	Edit: Volume is too low in noi	The volume level for me to he	1	sound
26	Amz US	1	Edit: Volume is too low in noi	For something that the compa	1	sound
27	Amz US	1	Edit: Volume is too low in noi	Two things I definitely don't li	1	sound
28	Amz US	1	Edit: Volume is too low in noi	The micro connector is hard to	1	design
29	Amz US	1	Edit: Volume is too low in noi	If I am bouncing in a machine	1	connectivity
30	Amz US	1	Edit: Volume is too low in noi	The rubber cover on the charg	1	power
31	Amz US	1	Doesnt live up to performance	I have had this product for jus	1	power
32	Amz US	1	Doesnt live up to performance	The volume level is also very	1	sound

Figure 20 - Screenshot of final dataset

5.1.4 Initial assessment from the dataset

5.1.4.1 Top products with most comments from consumers

Based on the new dataset, we can perform the initial analysis in terms of products with most comments from consumers. As displayed in figure 21, Jabra manager can instantly understand the top 10 Jabra products with most comments from consumers. The fact that consumers often discuss about these products has shown their popularity among the community. Thus, we recommend that Jabra needs to pay attention to the reviews related to these Product IDs to identify and improve these products in order to win their consumers.

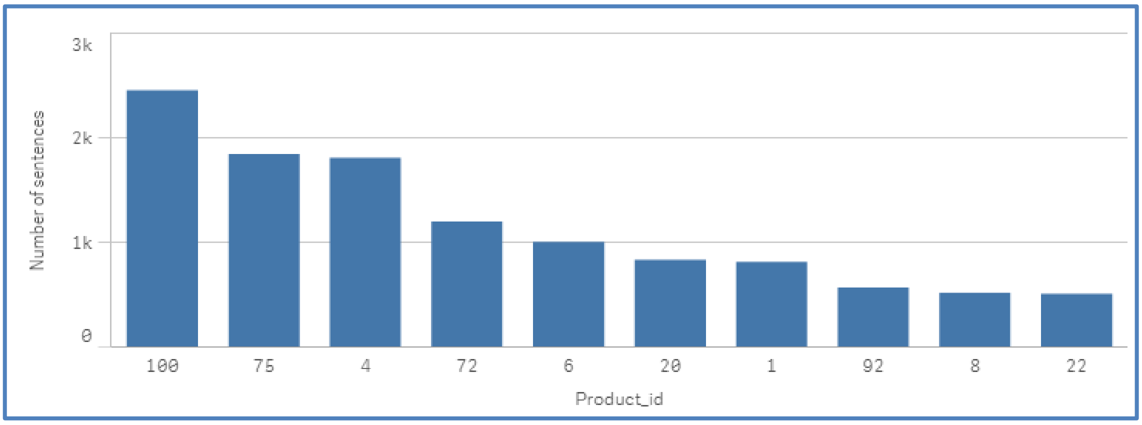


Figure 21 - Top 10 products with most comments

5.1.5 Result comparison with review approach

The purpose of this experiment aims to analyze customer reviews, extract their opinions and provide product quality assessment using text analytic techniques in an automatic way. Consequently, we intend to perform a comparison between these two approaches so as to confirm that our recommendation can deliver business values to the case company.

Firstly, we aim to evaluate the accuracy of our approach by comparing the result of feature tagging experiment with the key point field labelled by Jabra. We summarize the number of key points that were classified in two approaches:

- Review approach: the key point is identified from the review and classified manually by Jabra personnel.
- Sentence approach: the key point is identified on sentence level, and conducted automatically by Python program.

The percentages of key points identified from consumer reviews are displayed in the two (02) following pie charts.

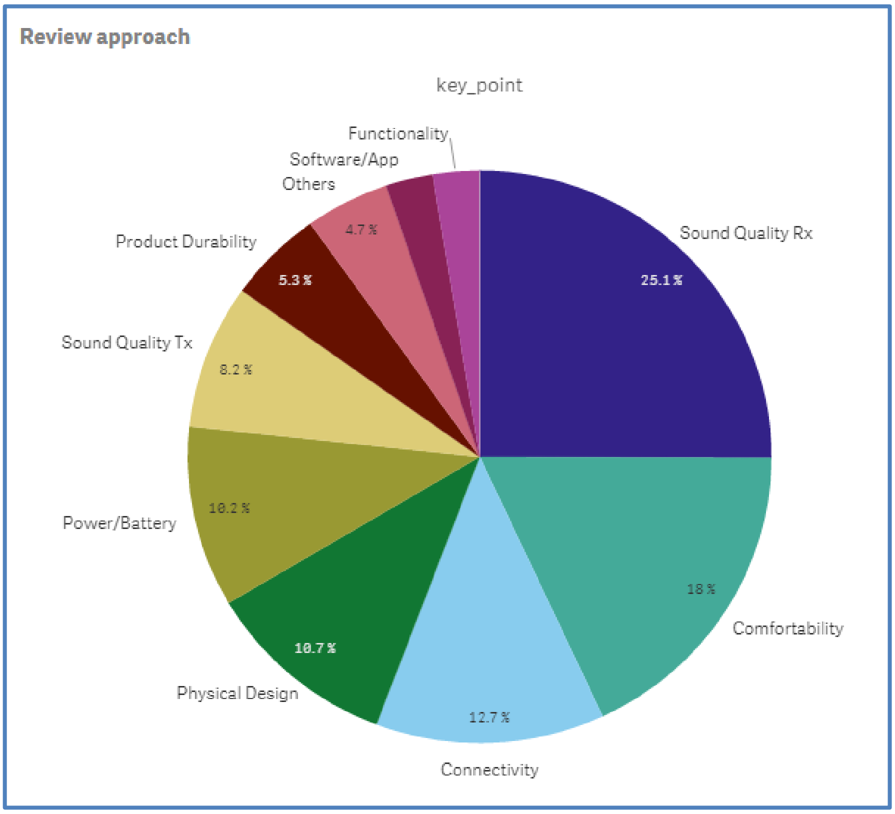


Figure 22 - Pie chart of key_point based on Jabra manual classification

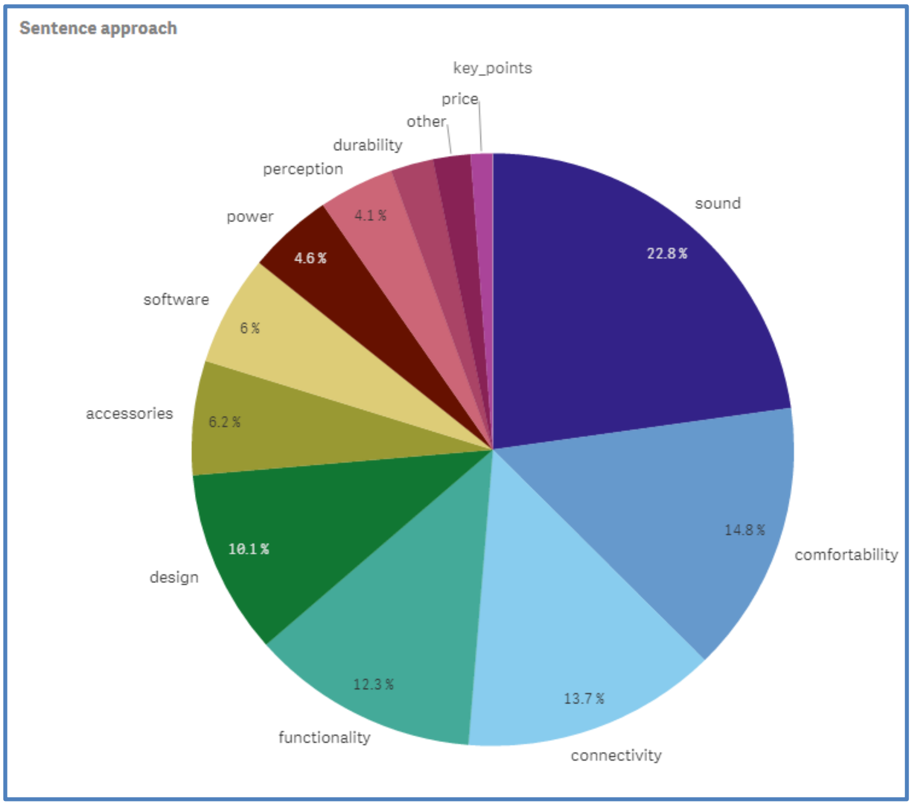


Figure 23 - Pie chart of key_point based on sentence approach

According to the pie charts, we noticed that the largest percentage of reviews were discussing about the sound quality (with 25.1% for Sound Quality Rx and 8.2% Sound Quality Tx from Review approach). The similar trend is observed from the Sentence approach (22.8%, correspondingly). The second most attentive feature of Jabra products is comfortability, with 18% in review approach and 14.8% in sentence approach. Followed is connectivity, with 12.7% and 13.7% for review and sentence approaches, respectively. Although there are some slight differences between these two patterns, they have shown that the automatic approach shows similar results to the manual method. The differences in sentence approach can be due to the reason that the provided list of key words is not exactly identical to the parameters used by manual approach. For instance, the categories used for sentence approach consists of categories which cannot be found in review approach, such as “perception”, “price” and “accessories”.

We also evaluate the accuracy of feature tagging experiment by analyzing features of Jabra products in term of rating. We summarize the results as in the following charts:

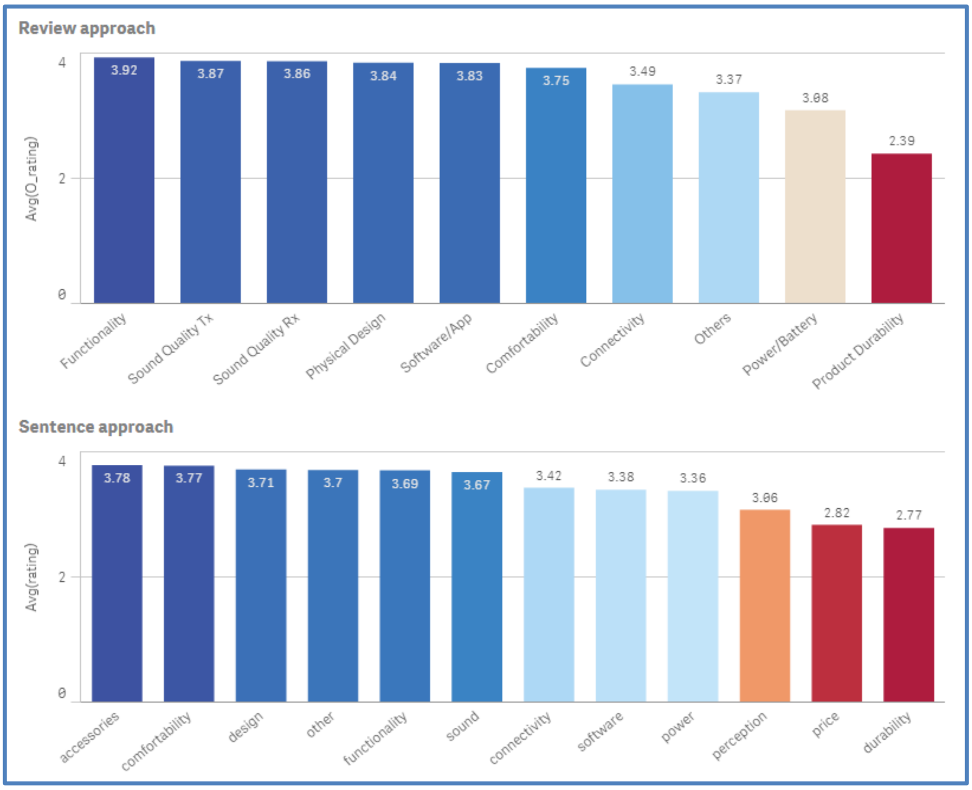


Figure 24 - Bar charts presenting comparison between review and sentence approaches

Both charts show that "Durability" is the weakest characteristic of Jabra products, which receives the lowest rating from consumers. Based on the review approach, consumers give only 2.39 out of 5 in average to this feature. On sentence level, it also receives a lowest rating (2.77 out of 5, in average). Since "perception" and "price" are not included in the original classification, we are not able to compare their ratings. From the original data, we realize that "Power" also receives low ratings from consumers, which is also reflected in the result from sentence approach (3.36 out of 5, which standing in the 4th at the bottom). Although other features are not ranked similar in these charts, the values may vary in minimal range (3.6-3.9), thus it has shown that sentence approach is appropriate, and provides acceptable result compared to the human assessment.

5.2 Experiment 1: Sentiment analysis

Sentiment analysis or opinion mining is the process of determining the attitude of a writer with respect to a topic. We need to consider the problem that reviews span over multiple sentences, while dealing with opinion mining on reviews. There are cases when a review contains multiple sentences and among them few sentences have opposite sentiment. For example, "This headset was superb, good quality and battery. The material it's made of was awful". In this review the first sentence shows positive polarity and the second sentence shows negative polarity. In order to avoid such problem, in our research, we applied sentiment analysis on the sentence-level, which means that sentence opinion towards one or more subjects is analyzed. It results in the overview of the sentence and explores information about subjectivity and bias of the sentence.

5.2.1 Sentiment analysis result & analysis

5.2.1.1 Overview of sentiment result

There are 17,314 records in total after data preprocessing phase. After running Python code to perform sentiment analysis on the newly created dataset, we summarize the result of sentiment values based on the labels as in figure 25. From the result, it is remarked that the majority of comments are labelled as "Positive" (62.74%). The "Neutral" comments are approximately equivalent to the "Negative" comments

(18.67% and 18.59%). The result of sentiment analysis is attached in Appendix D.

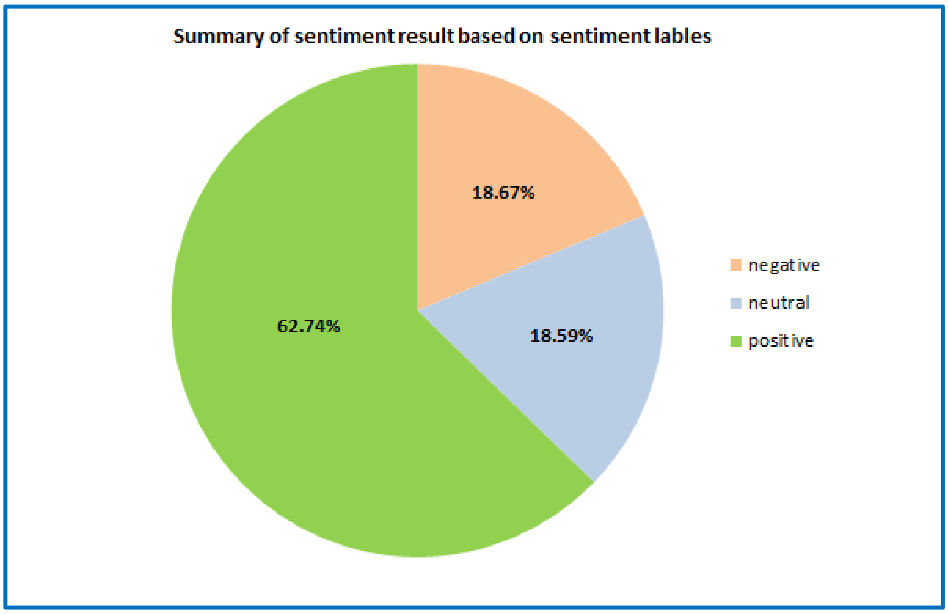


Figure 25 - Pie chart summarizing sentiment analysis

When calculating the average of polarity values, we notice that the average polarity of "Positive" records has higher absolute value compared to those of "Negative" records, which can draw a conclusion that the majority of Jabra consumers are delighted in their products (refer to table 3). Moreover, the absolute value of average polarity for positive sentences is higher than the one for negative sentences, this leads us to believe that when customers write a positive review (sentence level), they use richer positive expressions and upraise adjectives, whereas, when writing negative reviews the expressions are not strong or rich comparatively.

Row Labels	Average of Polarity
negative	-0.2668125
neutral	0
positive	0.38391861
Grand Total	**0.157904042**

Table 3- Table summarizing average of polarity values

Based on the pie chart and the table of average polarity values, Jabra quality manager can have an overall assessment about their products in the market. It can suggest that most of Jabra consumers are satisfied with their experience with Jabra products. Also, the small percentage of "Neutral" reviews can be explained by the fact that the dataset does not include irrelevant data, unlike dataset retrieved from other social media platforms. For instance, comments retrieved from Facebook usually contains auto-comments, spamming or advertisements, which are usually classified as "Neutral", but they are irrelevant and do not help analysts in the product assessment. Therefore, it is also supportive in proving the fact that the dataset is relevant to our research, so it can be applied to further analysis in other experiments.

5.2.1.2 Top products based on sentiment result

Another application of sentiment analysis in product quality assessment is the display of top products receiving most concerns from the public. Figure 26 displays the top 10 Jabra products with most positive comments from consumers, while figure 27 displays the top 10 Jabra products with most negative comments.

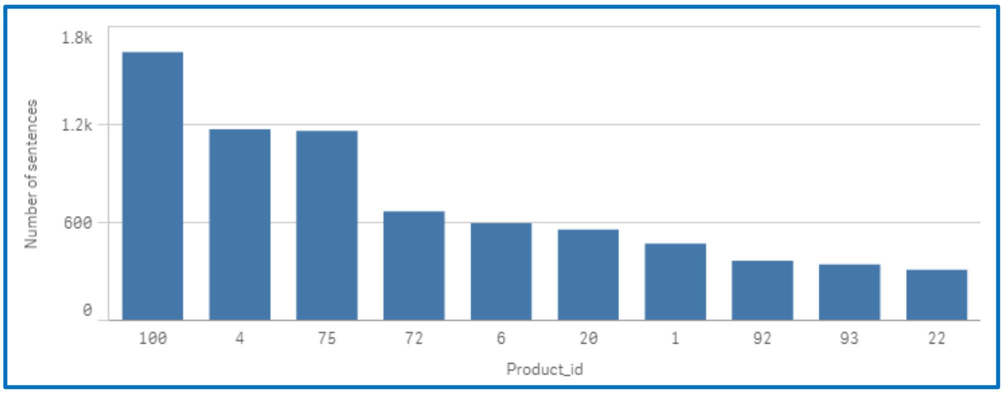


Figure 26 - Top 10 products with most positive comments

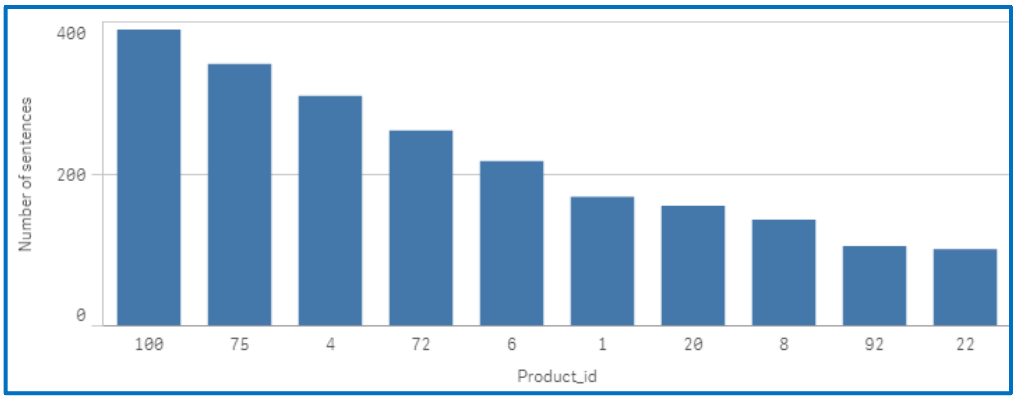
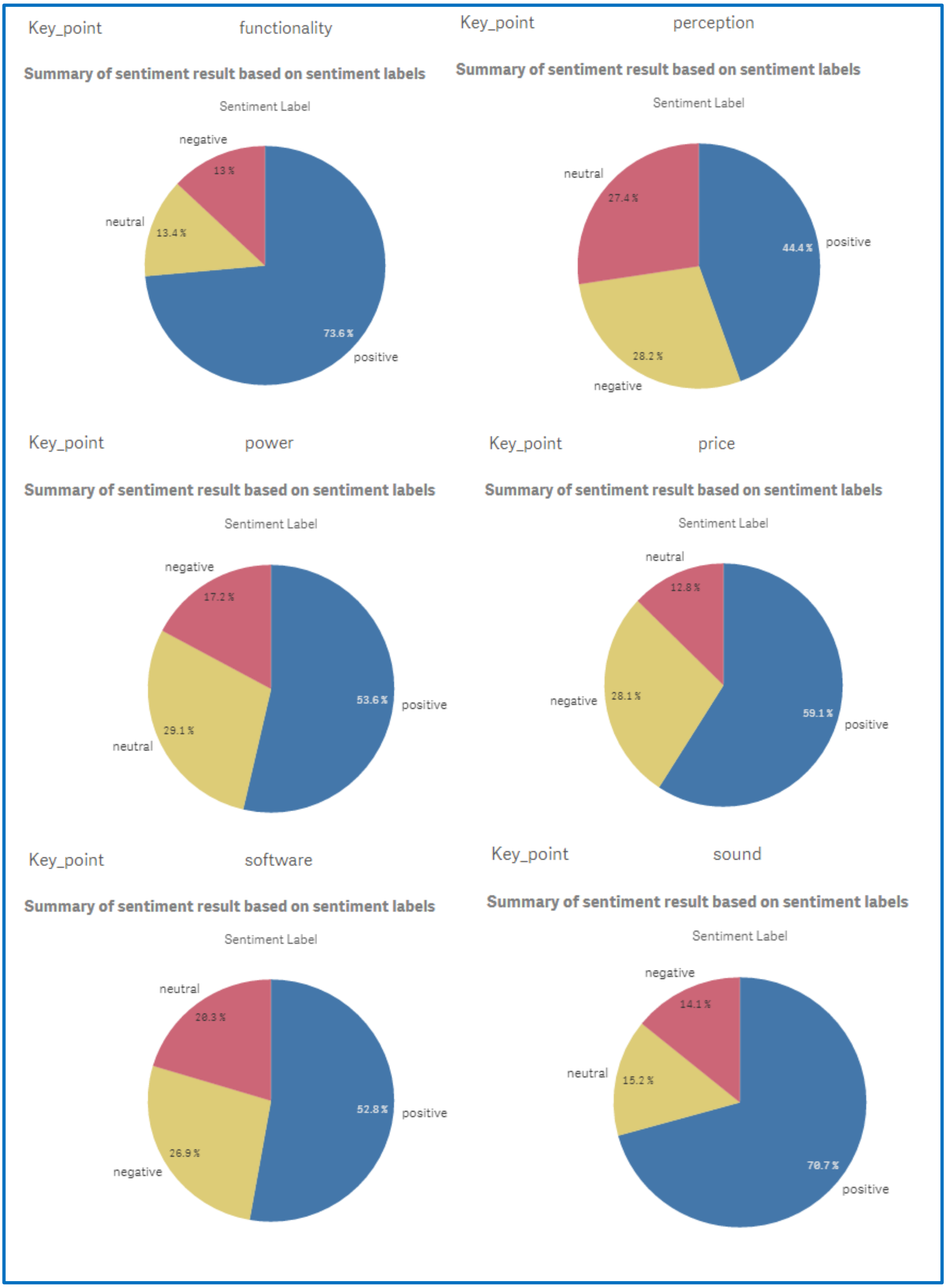


Figure 27 - Top 10 products with most negative comments

These graphs have shown that these top 10 products are the most popular Jabra products among the consumer community, receiving most comments from consumers, both positive and negative. It implies the fact that if Jabra invest in improving these products, they can increase customer satisfaction.

5.2.1.3 Sentiment result on feature-level

The sentiment result can also be assessed in feature - level in order to see how consumers reflect their experience with Jabra products in term of product features. When addressing the sentiment values for specific features, the sentiment result for each feature is displayed in the following charts:


Key_point functionality
Summary of sentiment result based on sentiment labels
Sentiment Label
negative
13 %
neutral
13.4 %
73.6 %
positive
Key_point perception
Summary of sentiment result based on sentiment labels
Sentiment Label
neutral
27.4 %
44.4 %
positive
28.2 %
negative
Key_point power
Summary of sentiment result based on sentiment labels
Sentiment Label
negative
17.2 %
53.6 %
positive
29.1 %
neutral
Key_point price
Summary of sentiment result based on sentiment labels
Sentiment Label
neutral
12.8 %
negative
28.1 %
59.1 %
positive
Key_point software
Summary of sentiment result based on sentiment labels
Sentiment Label
neutral
20.3 %
52.8 %
positive
26.9 %
negative
Key_point sound
Summary of sentiment result based on sentiment labels
Sentiment Label
negative
14.1 %
neutral
15.2 %
70.7 %
positive

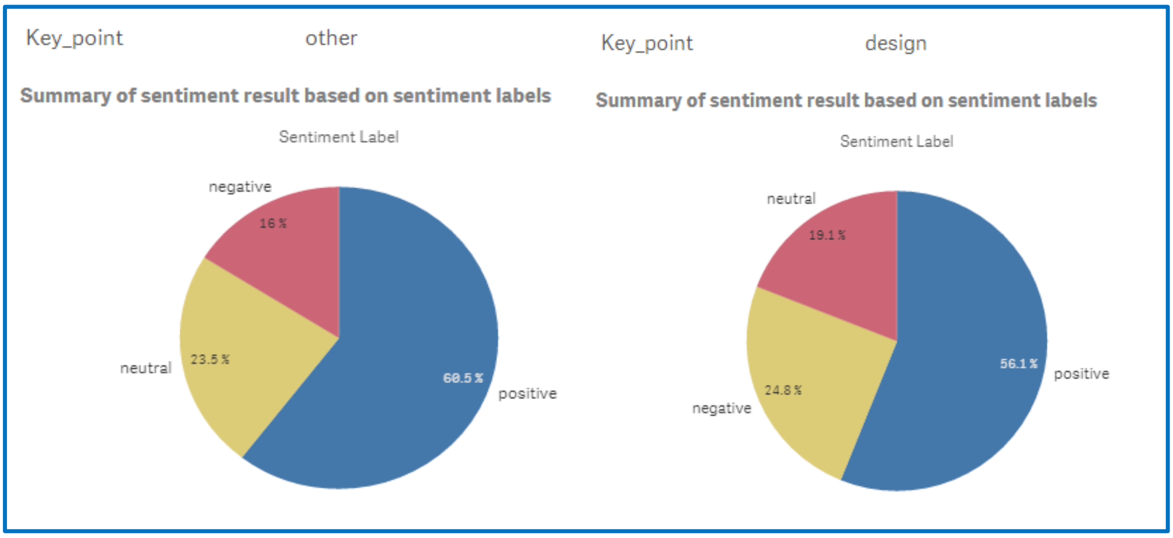


Figure 28 - Summary of Sentiment analysis for each key_point

All of these charts are sharing a similar trend, where most of the sentences are classified as "Positive". However, there are two attributes which receive the "Positive" feedback less than 50% of the total reviews, which are "Durability" (44.2%) and "Perception" (44.4%). From these charts, Jabra product manager can have an overview about the satisfaction of their consumers, and what they are complaining about so as to have plans for product improvement in the future.

5.2.2 Application of sentiment analysis in BI tool

5.2.2.1 Detection of product with weak points

During the first experiment we run the sentiment analysis of the dataset using Textblob in Python. The rationale was to identify positive, negative and neutral sentiment or opinion in these reviews. The positive reviews are powerful indicators of what consumers like about the product. On the other hand, negative sentiment shows what product areas need improvement. The results are quite valuable for the quality management team of Jabra. The result of sentiment analysis can be further applied to assess product quality. For example, based on the rating provided by consumers, the company can specify that "durability" is the weakest feature in Jabra products, which the average rating is only 2.77/5.0:

Figure 29 - Dashboard sample of weak point detection

This feature is further analyzed in term of polarity values retrieved from sentiment analysis, and analysts can identify specific products receiving negative reviews from consumers regarding "durability" feature. From the chart, products with polarity values less than 0 have received negative feedback from consumers, and the result is translated that the smaller the polarity, the more negative feedback this product was given. Products received most negative feedback about "durability" are products with id 55, 19, 52 or 63. From this list, Jabra can perform more detailed assessment over defined products, so as to detect weak points and improve the products.

5.2.1.2 Overview for product quality assessment

Understanding the importance of getting to know the consumer feedback about products, this section aims to provide an effective method to assess a product ID. By importing the sentiment result and using BI tool to slice-and-dice the data, analysts can provide a fast appraisal on the product. Figure 30 is the demonstration of how to leverage the sentiment results in understanding consumer' opinions and product quality assessment.

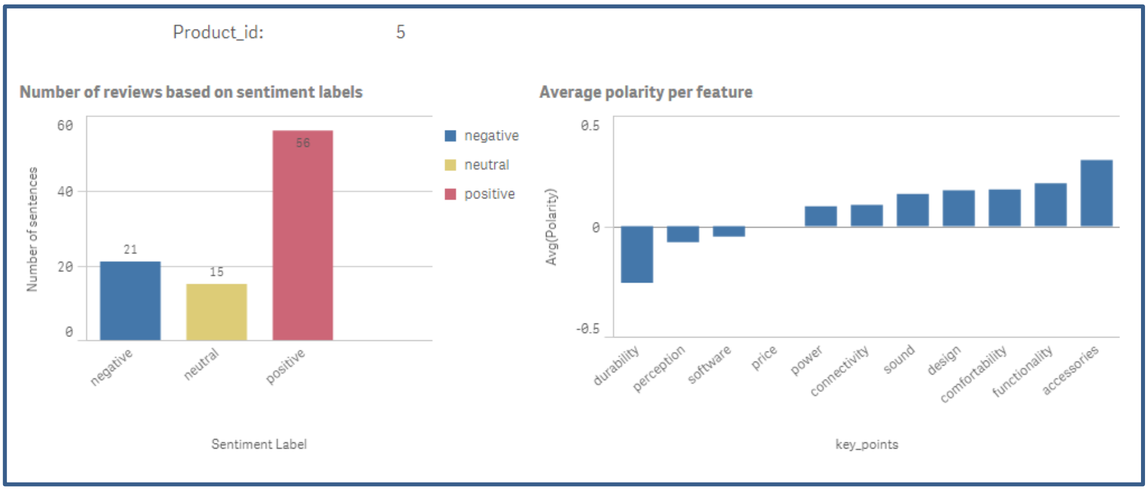


Figure 30 - Dashboard sample of product quality assessment

The above figure demonstrated the sentiment value of product_id 5. From the chart, analyst can extract critical factors, such as:

- There are 56 positive sentences, 15 neutral sentences and 21 negative sentences. The majority of reviewers were expressing good experience with this product.
- The product has both favorable and unfavorable features. Some features have high score, showing the satisfaction of consumers, such as "accessories", "functionality" or "comfortability". Some features, on the other hand, have negative score, which is a sign of dissatisfaction of consumers, such as "durability" (-0.26), "perception" (-0.07) or "software" (-0.05). Feature "price" has the average polarity as 0, implying that consumers have neutral opinion about the product price. Consequently, Jabra product quality manager should perform closer analysis on products receiving low scores from consumers.

5.2.3 Result comparison with review approach

We summarize the positive and negative reviews per product features. Based on Jabra classification, there are two types of feedback, which are 1 (positive) and 0 (negative). Per observation, almost all features of Jabra products received positive feedback from consumers, except for "durability".

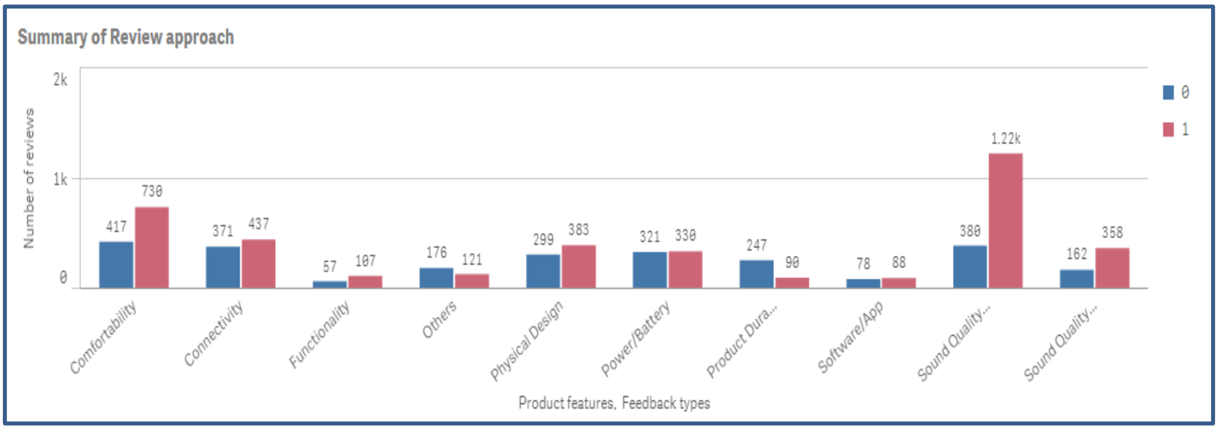


Figure 31 - Summary of manual sentiment classification

The scope of this study is not to compare results from manually classified data with sentiment classification results; however, we found similar patterns between the two approaches. It is also noteworthy that manually labeled data consist of only positive and negative labels, and sentiment classification result shows 03 (three) types of labels including positive negative and neutral. Thus, the comparison is somewhat unjust; however, for the sake of aligning our results with that of manual process, we measure the two dataset result responses.

For sentence approach, we use the sentiment result for classification, and the result is displayed in the following bar chart, showing that the majority of reviews are also classified as "Positive".

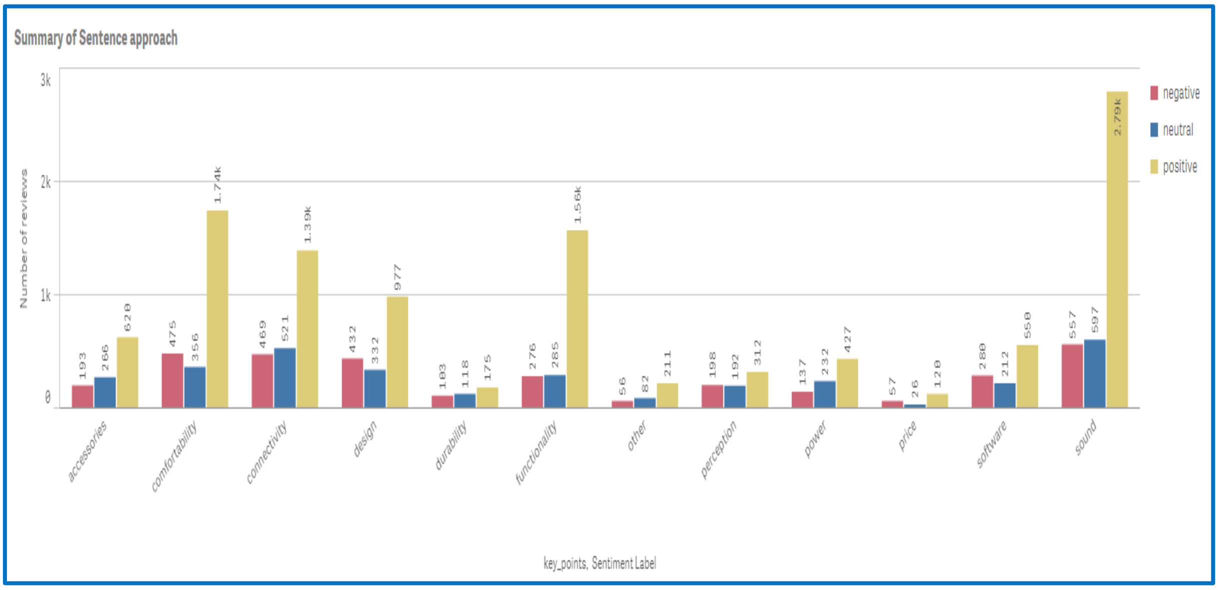


Figure 32 - Summary of automatic sentiment analysis

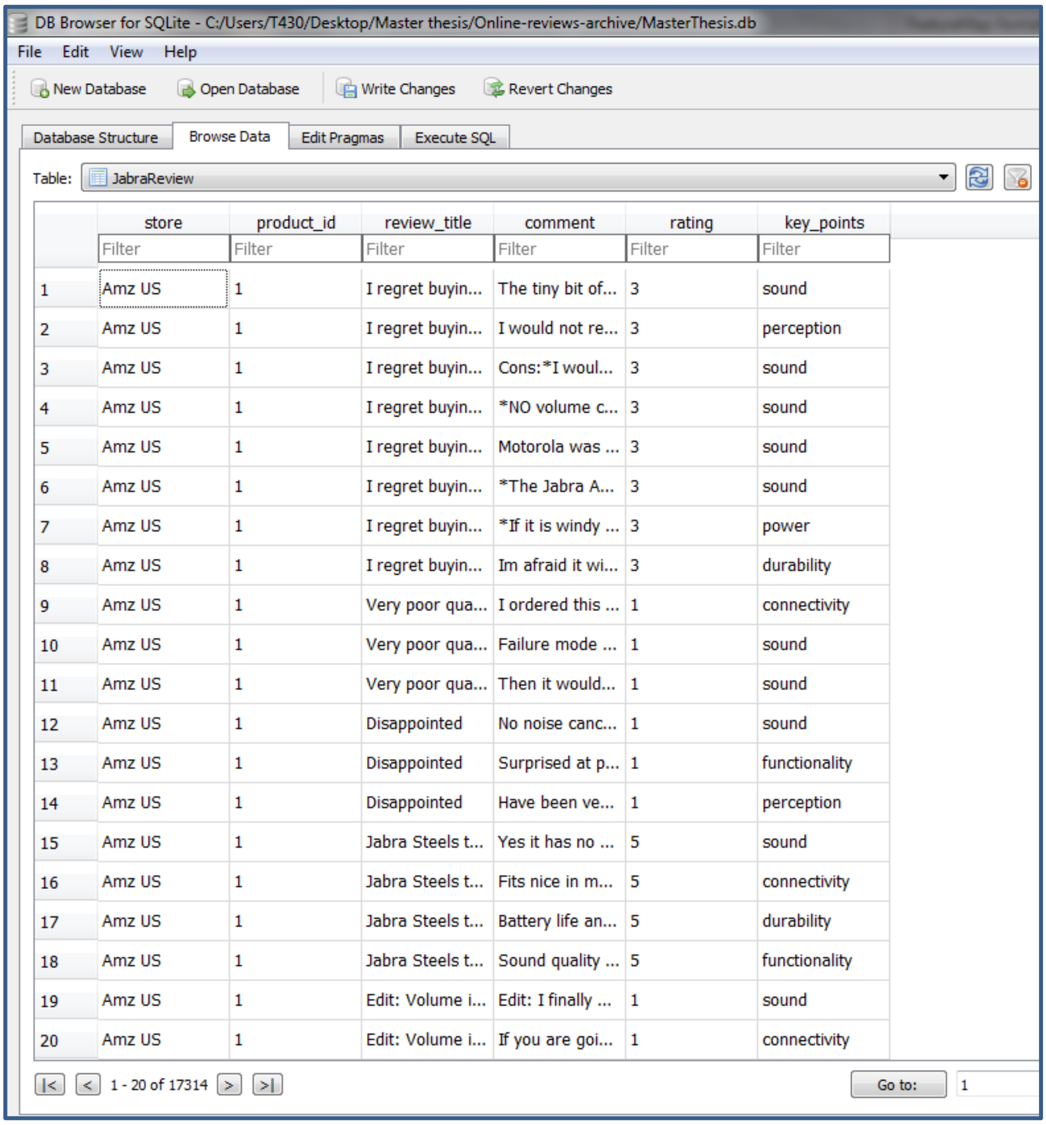
5.3 Experiment 2: OLS Regression analysis

The purpose of this experiment aims to answer the second sub-research question to suggest a method of automatic extraction of descriptive words from the consumer feedback. During the second experiment, the dataset is used in OLS regression analysis to automatically identify tokens (as independent variables) with features (dependent variables). These results help identify key words occurring in customer reviews every time they mention “comfortability” or “sound” and the other 10 performance attributes developed by Jabra. The idea behind this technique is that real product features are likely to be commented on by many customers.

Additionally, the POS tagging for the tokens are labelled. It provides further insights into these descriptive words and supports analysts in coming up with more meaningful analysis results.

5.3.2 Database creation

Figure 33 describes the structure of table in SQLite. The total number of records after feature mapping is 17,314 records.



	store	product_id	review_title	comment	rating	key_points
	Filter	Filter	Filter	Filter	Filter	Filter
1	Amz US	1	I regret buyin...	The tiny bit of...	3	sound
2	Amz US	1	I regret buyin...	I would not re...	3	perception
3	Amz US	1	I regret buyin...	Cons:*I woul...	3	sound
4	Amz US	1	I regret buyin...	*NO volume c...	3	sound
5	Amz US	1	I regret buyin...	Motorola was ...	3	sound
6	Amz US	1	I regret buyin...	*The Jabra A...	3	sound
7	Amz US	1	I regret buyin...	*If it is windy ...	3	power
8	Amz US	1	I regret buyin...	Im afraid it wi...	3	durability
9	Amz US	1	Very poor qua...	I ordered this ...	1	connectivity
10	Amz US	1	Very poor qua...	Failure mode ...	1	sound
11	Amz US	1	Very poor qua...	Then it would...	1	sound
12	Amz US	1	Disappointed	No noise canc...	1	sound
13	Amz US	1	Disappointed	Surprised at p...	1	functionality
14	Amz US	1	Disappointed	Have been ve...	1	perception
15	Amz US	1	Jabra Steels t...	Yes it has no ...	5	sound
16	Amz US	1	Jabra Steels t...	Fits nice in m...	5	connectivity
17	Amz US	1	Jabra Steels t...	Battery life an...	5	durability
18	Amz US	1	Jabra Steels t...	Sound quality ...	5	functionality
19	Amz US	1	Edit: Volume i...	Edit: I finally ...	1	sound
20	Amz US	1	Edit: Volume i...	If you are goi...	1	connectivity

Figure 33 - Database structure in SQLite

5.3.3 Tokenization & Stop-word removal

In order to demonstrate these two steps, we will analyze one sentence from the dataset as a sample.

Sentence: “fits nice in my ear but I can see people with small ear canals possibly being discomforted if they use the provided ear hook pictured."

The result of tokenization step includes a list of tokens generated by splitting this sentence into single terms, which is as followed:

"fits", "nice", "in", "my", "ear", "but" ,"I", "can", "see", "people", "with", "small", "ear", "canals", "possibly", "being", "discomforted", "if", "they", "use", "the", "provided", "ear", "hook" ,"pictured".

Afterwards, the stop word removals step will help remove all English stop words from the tokenized sentences. Consequently, the remaining list of tokens is as follows:

"fits", "nice", "ear", "see", "people", "small", "ear", "canals", "possibly", "discomforted", "use", "provided", "ear", "hooked" and "pictured".

In the implementation, Tokenization and Stop word removal are the two steps in TF-IDF calculation, which are embedded into the operation of TfidfVectorizer. Therefore, results from tokenization and Stop words removal are not explicitly displayed but further used in TF-IDF calculation.

5.3.4 OLS regression in extracting descriptive tokens

Generally, there are words or phrases which are more helpful in describing a specific phenomenon compared to the rest of words in sentences. Consequently, this experiment leverages techniques in data analytics to find distinctive tokens of known attributes of Jabra products (as defined by Jabra expert).

The prerequisite for OLS regression experiment is the tokens with TF-IDF values, which were pre-processed and calculated as previously explained in section 5.1. Because the primary goal of this step is to obtain most distinguishing tokens for each product attributes, those tokens are examined in the context of the entire corpus. The result of this experiment shows the strength of a token in relation to an attribute.

The following section is the demonstration of applying OLS regression analysis in measuring the relationship between token and one product attribute: "*durability*". The reason drives us to select this attribute for demonstration as it is the attribute receiving the most negative reviews from consumers.

From the list of tokens created, each token will be considered as the independent variable, with its TF-IDF value as the variable value. The dependent variable has 02 (two) values: 0 if no existence of feature, and 1 if there is feature existence in the sentence. By running OLS for the word in the entire corpus, it returns the OLS result. For instance, the following figure demonstrates the OLS summary result for the word “*warranty*” in predicting the attribute “*durability*”.

```
Word: warranty
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.128
Model:                            OLS   Adj. R-squared:                  0.127
Method:                 Least Squares   F-statistic:                     2530.
Date:                Thu, 27 Apr 2017   Prob (F-statistic):               0.00
Time:                        23:04:44   Log-Likelihood:                 9318.3
No. Observations:               17314   AIC:                        -1.863e+04
Df Residuals:                   17313   BIC:                        -1.863e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
x1             0.7554      0.015     50.302      0.000         0.726     0.785
==============================================================================
Omnibus:                    19235.979   Durbin-Watson:                   1.868
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1481868.007
Skew:                           5.847   Prob(JB):                         0.00
Kurtosis:                      46.788   Cond. No.                         1.00
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 34 - Sample of OLS result summary for one word

From the result of OLS regression, there are several important elements in interpreting the relationship between these variables:

- R-squared, or Coefficient of Determination: it is a statistical measure of how well the regression line represents the data. The value of R-squared as 0.128 which means that 12.8% of the total variation in y can be explained by the linear relationship between x and y (as described by the regression equation).
- coef : it is the estimation of coefficient. In case of the token “warranty”, the value is 0.7554. In the research, this element is used to measure the strength of correlation between two variables.
- P > |t|: it is the p-value of the null-hypothesis that the coefficient = 0 is true. Regularly, the relationship between the independent variable and dependent variable is considered as statistically significant when p-value less than the

confidence level. The confidence level is often set as 0.05, so only words with p-value <0.05 is considered for analysis.

By running OLS regression for all tokens defined previously, and filter those with statistically significant correlation, we can develop list of tokens which are highly related to the product feature. The following table demonstrates the OLS regression for one product attribute - "*durability*".

Token	Coefficient	p-value
warranty	0.755373767	0
hot	0.735353908	3.43E-101
damaged	0.674066632	7.48E-25
break	0.635031039	7.81E-246
breaking	0.556963337	7.62E-102
broken	0.53006402	2.56E-140
weather	0.50533236	2.72E-16
breaks	0.49481455	1.15E-55
damage	0.485448607	1.57E-38
resistant	0.481950615	4.07E-95
waterproof	0.478874346	9.93E-143
breaker	0.454941446	1.99E-41
period	0.449764221	2.22E-39
defective	0.42153178	7.23E-143
proof	0.41907649	5.53E-30
pool	0.40828534	1.09E-13
photo	0.397176298	1.97E-36
defect	0.379992477	4.48E-19
contacted	0.359339448	2.00E-14
water	0.355861631	1.05E-46
sweat	0.353564305	4.10E-67
original	0.352311904	3.13E-13
incase	0.352116154	5.48E-08
jawbone	0.349323355	1.47E-06
130	0.340508242	0.004059158
year	0.293819315	6.62E-48
essentially	0.290590526	0.001251039
covered	0.286690292	1.30E-19
deal	0.276384764	6.63E-27
junk	0.249099683	5.77E-08

Figure 35 - Screenshot from table of OLS result for "durability"

These are the top 30 tokens which have the strong correlation with the topic "*durability*" (measured by the coefficient values). The coefficient of the target explanatory variable is taken as the ***strength of relationship*** (Schwartz et al., 2013). The higher the coefficient value is, the stronger the correlation between the token and the topic "*durability*". These words well describe characteristics of the "*durability*" feature, such as related components like "waterproof", "resistant" or "warranty". Also, words like "pool" or "water" can help analysts make an assumption that consumers are usually using their products for fitness activities like swimming etc. Thus it can be assumed that consumers are more interested in the quality of product in the condition of using for long time or being waterproof. Comparing to those words defined by Jabra, there are some descriptive words which are interesting to analyze, for instance, "photo" or "jawbone". They are not usually linked together with "*durability*" in the common perception, but the result from OLS regression has shown their strong correlation with the "*durability*" attribute of Jabra products. Hence, the list of tokens can help analysts to have different insights into their products, as well as refine their feature classifications based on keywords.

The OLS regression analysis for all feature attributes is attached in Appendix E. The result provides significant insights, when comparing OLS regression result with the manual performance indicators. This might help us explain how automatic processes, when it comes to large amount of data perform better when it comes to identifying performance of products and attributes.

There are some differences in terms of correlation strength among topics. As demonstrated in the figure 36, attribute "sound" has the highest average coefficient, which means that words identified in regarding to "sound" has a relatively strong correlation with this attribute. It can be explained by the fact that consumers tend to use a more convergent set of vocabularies when discussing about the attribute "sound". On the other hand, attribute "price" has the lowest average coefficient, showing a loose correlation between the topic and words. It can be explained by the fact that consumers can have various ways to discuss about this topic, either directly ("cheap" or "expensive") or indirectly ("great deal", "big save", ect.). Secondly, some words can be used in more than one situation (for instance, "cheap" can also be used as "cheap looking", "cheaply built" or "cheap quality"). Therefore, these words have correlated to several attributes rather than "price", and it reflects in the lower correlation strength.

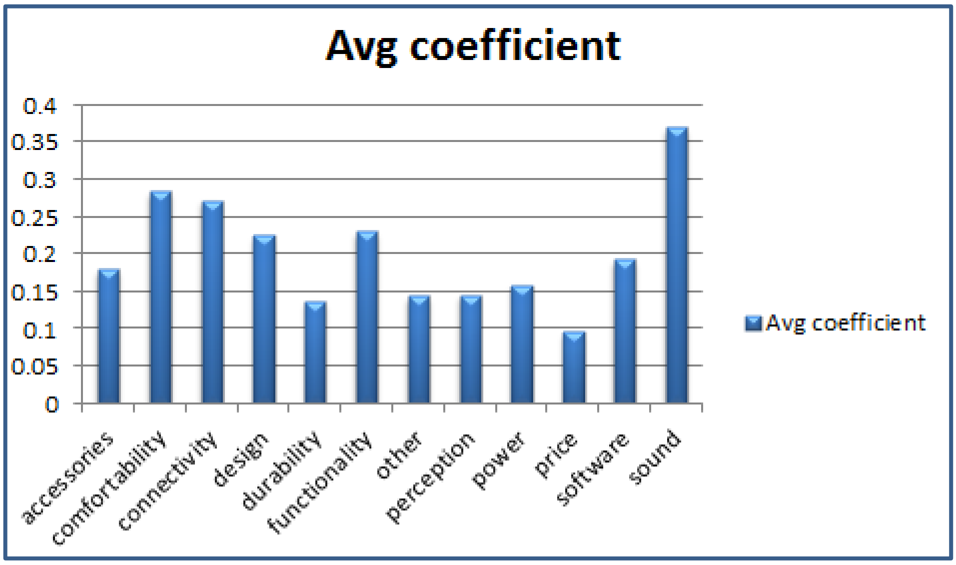


Figure 36 - Bar chart of average coefficient per key_point

The result from average coefficients has shown that the automatic approach for feature categorization will perform differently on topics. For topics with high average coefficients, this approach will give better results since consumers tend to use more explicit words. Attributes with low average coefficients show that consumers use more ambiguous words and can carry other latent meanings.

5.3.5 OLS regression in extracting descriptive tokens from negative reviews

In relation to product quality assessment, analysts highly concern about products with negative feedback from consumers, because they can reveals aspects for improvement. In order to fulfill the further requirements, the experiment can be extended to include rating given to reviews by consumers. It is assumed in the context of experiment that 1 star rating ("*I hate it*") provides most of negative feedback for products.

We continue with detecting words with highly correlated to attribute "*durability*" in the context of negative reviews, as the extension of the previous example. By applying similar techniques, we develop list of words highly correlated with "*durability*" in negative reviews (figure 37).

Token	Coefficient	p-value
warranty	0.421618208	0
broken	0.378927775	1.36E-203
130	0.340508242	1.34E-06
defect	0.317480695	3.76E-36
defective	0.263013886	2.03E-157
period	0.262890121	4.16E-38
junk	0.249099683	6.71E-20
covered	0.240143813	2.14E-37
selling	0.234261576	4.25E-12
exposing	0.234071505	2.94E-13
responsive	0.228452761	1.95E-16
break	0.214538376	5.37E-79
claimed	0.212290236	2.15E-13
exact	0.210956523	2.98E-06
breaks	0.210372459	3.29E-29
contacted	0.19782182	1.41E-12
email	0.186426377	8.42E-11
priced	0.177621674	3.43E-11
process	0.170253052	3.65E-29
60	0.163787926	7.69E-05
center	0.161347129	2.03E-08
popping	0.160947211	0.001407
told	0.159549752	2.27E-22
seller	0.158602459	5.68E-13
breaking	0.157881482	2.91E-24
customer	0.157216283	7.47E-25
replacement	0.156580643	4.99E-39
damage	0.145942674	5.69E-11
contact	0.142772337	2.05E-11
locate	0.142640402	1.50E-06

Figure 37 - Screenshot from table of OLS result for "durability" in negative comments

The results provide significant insight for Jabra to evaluate their product quality. From the top words highly correlated with "*durability",* it is remarkable that consumers describe their experience with Jabra product as "broken" or "defective". This reveals a weak point of Jabra product which managers should pay attention towards. Additionally, the list of words can suggest some ideas for Jabra managers. For instance, the appearance of words like "email", "responsive" or "contact" reveals customer's attention to the after-sale customer service. Therefore, the organization needs to consider the appearance of word that are out of context such as mentioned above, as that might carry important piece of information regarding their consumer support.

5.3.6 POS tagging

The objective of POS tagging is to identify the part of speech of the token, and label it appropriately. By grouping tokens into their POS categories, analysts can perform further review in specific groups of tokens. For instance, in regarding to the list of tokens generated for attribute "*durability*" in negative feedback, the list of tokens with their corresponding POS tag is built as in figure 38. By filtering only adjectives (POS tag = JJ or JJR (comparative adjectives) or JJS (superlative adjectives)), we get the list of adjectives with high correlation with "*durability*" (figure 39).

Token	Coefficient	p-value	POS tag
warranty	0.421618208	0	NN
broken	0.378927775	1.36E-203	NN
130	0.340508242	1.34E-06	CD
defect	0.317480695	3.76E-36	NN
defective	0.263013886	2.03E-157	JJ
period	0.262890121	4.16E-38	NN
junk	0.249099683	6.71E-20	NN
covered	0.240143813	2.14E-37	VBN
selling	0.234261576	4.25E-12	VBG
exposing	0.234071505	2.94E-13	VBG
responsive	0.228452761	1.95E-16	JJ
break	0.214538376	5.37E-79	NN
claimed	0.212290236	2.15E-13	NN
exact	0.210956523	2.98E-06	NN
breaks	0.210372459	3.29E-29	NNS

Figure 38 - Screenshot from table of OLS result with POS tag

Token	Coefficient	p-value	POS t
defective	0.263013886	2.03E-157	JJ
responsive	0.228452761	1.95E-16	JJ
original	0.138372808	1.49E-06	JJ
resistant	0.124743783	3.17E-19	JJ
live	0.090343363	0.001172379	JJ
hot	0.080880935	8.65E-05	JJ
likely	0.079668033	0.006271687	JJ
waterproof	0.075008811	3.07E-11	JJ
crazy	0.072453446	0.011906719	JJ
careful	0.068586717	0.003200061	JJ
new	0.047227123	1.65E-05	JJ
second	0.035733026	6.50E-05	JJ
free	0.035249282	0.048753756	JJ
terrible	0.029744566	0.000233644	JJ

Figure 39 - Screenshot from table of OLS result with POS tag as Adjectives

The list of adjectives regarding to the attribute "*durability*" is helpful for analysts, as it explicitly explains their customer satisfaction level with products. Terms such as "resistant" or "waterproof" may describe characteristic of product feature, while terms like "new" or "free" may express customer expectation for products. Lastly, words such as "crazy" or "terrible" are helpful in showing level of customer feelings about products.

5.3.7 Data visualization

In order to intuitively summarizing results from OLS regression analysis, we use word clouds. Nonetheless, unlike most of word clouds which use word frequency to scale the word size, we use the coefficient value as the measurement for word size, that is larger words indicate stronger correlations with the topic. We also use different colors to represent different POS groups, which are blue for nouns, green for adjectives and orange for all verb forms. The remaining words which do not belong to these 03 (three) categories are grouped into one group.

The following figure presents word clouds for attribute "*durability*". Words are scaled based on the coefficient values from OLS regression analysis all reviews, regardless of their sentiment labels. They are effective in demonstrating how consumers express their opinions about Jabra products, and serve as a good tool of capturing consumer feedback. For instance, the word clouds for "*durability"* clearly show that this attribute has certain concerns from consumers. The most distinguishing words in verb cloud are "damaged", "break", "breaking" and "broken", showing what consumers are complaining about.

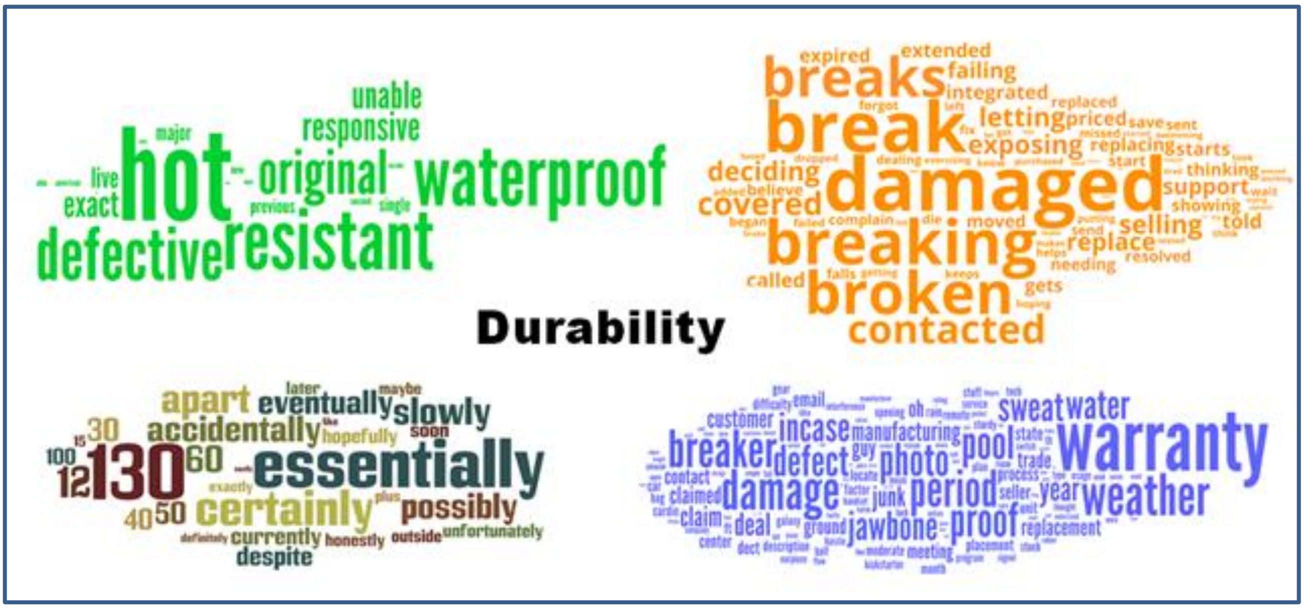


Figure 40 - Word clouds for "durability'

5.4 Valuable outcomes

5.4.1 Short term

The result of the research reveals the fact that Jabra products are receiving favorable reviews from the majority of users over the two e-commerce platforms, Amazon.com and Bestbuy. Also, the result supports the case company to understand top products which receive most attention from the public. Nevertheless, the result also shows that some products are still not able to entertain their consumers up to their expectation, demonstrated by the low average polarity that given to these products. Therefore, it provides the Jabra product quality manager an overview about how the market reacts to their products.

5.4.2 Mid term

Since the Internet has become ubiquitous in recent days, the buying habit of customers has changed. Prior to making decision to purchase a new product, customers prefer browsing for other users' reviews on the Internet. Therefore, the result of this research can also serve as a deciding factor for potential users prior making purchase decision. Also, from the company's perspective, they can utilize the results to get to know their consumers. Based on the results, Jabra should perform comprehensive assessment over products with most negative reviews from users.

5.4.3 Long term

The result of this dissertation can also be applied to observe how Jabra will respond to the market. The fact that Jabra can adjust a product based on consumer complaints or reviews can reflect how Jabra captures their consumer expectation. If consumers recognize that their opinions are valued and taken into consideration, their level of satisfaction will certainly increase. The appropriate response from Jabra will return in the high level of customer loyalty.

In addition, this research has confirmed the fact that social data does hold values for organization. Consequently, it suggests another implication for the case company, which is the new product development strategy. The automatic approach introduced in this dissertation can suggest a process-oriented approach for Jabra in capturing their consumer opinions and translating them into business values. We firmly believe that the case company can gain benefits from the introduced approach, as it is cost-effective, reliable and less manual error. Therefore, the case company can establish a new product development strategy to take advantages of this research.

CHAPTER 6

Discussion

6. Discussion

The experiments gave interesting insights about customer data reviews. In this chapter, we would like to discuss the findings of our results based on the data analysis done on consumer reviews and opinions about Jabra products from social platforms. In section 6.1, we would like to discuss the principal findings from our results in the light of the concepts that we discussed earlier in section 2.1 of conceptual framework. We also present the answers to research question in section 6.2. Section 6.3 and 6.4 will discuss the implications of our study. Section 6.5 presents our recommendations for the case company. In section 6.6, we admit the limitation of our research. Finally, in section 6.7, we discuss opportunities for future research based on this study.

6.1 Findings in light of concepts

6.1.1 Social media and two way communication

Social media platforms and consumer participation have strong influence not only on the way people communicate but also transforming the organization's' strategy on how to approach and manage loyal customer relationship. Jabra also realizes the importance of sentiment that customer reviews can carry on social media and therefore, is interested in finding what kind of opinion these reviews carry about their product quality. It is also important to note that the structure of platforms also plays an important role in writing and sharing review. For example, Amazon facilitates its customers with the ability to rate and write a review about the product. It also enables consumers to rate a review as how many people found the review helpful. This motivates customers to write reviews and potential buyers can make their purchase decisions based on these reviews.

Moreover, the biggest advantage these reviews hold is for Jabra itself as these reviews may help Jabra to improve their product quality and brand perception based on these reviews. This eliminates the extra effort of finding out whether customer likes the product or not through surveys or other marketing strategies. The e-commerce platforms have changed the traditional one way communication method of company sending marketing message to consumers, into this rich and interactive two-way communication process.

6.1.2 User generated content & Credibility of online customer reviews

For the sake of our research, we are not interested in the argument that whether consumer reviews hold more credibility than typical advertisements. However, it is clear that the shift in market trend has influenced the consumer behavior when it comes to making purchase decisions. Advertisements may have an influence on introducing new products in the market, but when it comes to decision making about a certain purchase, the potential customer wants to know the experience and reviews about the product. This does not require any prior knowledge about the person writing the review as websites or social media platforms adds credibility to the customer reviews. For example, customer reviews that are generated on Amazon and Bestbuy have registered customer profiles along with detailed review. This minimizes if not completely eliminates the factor of fake reviews. The following figure 41 is a screenshot taken from Amazon.com as an example of customer reviews.

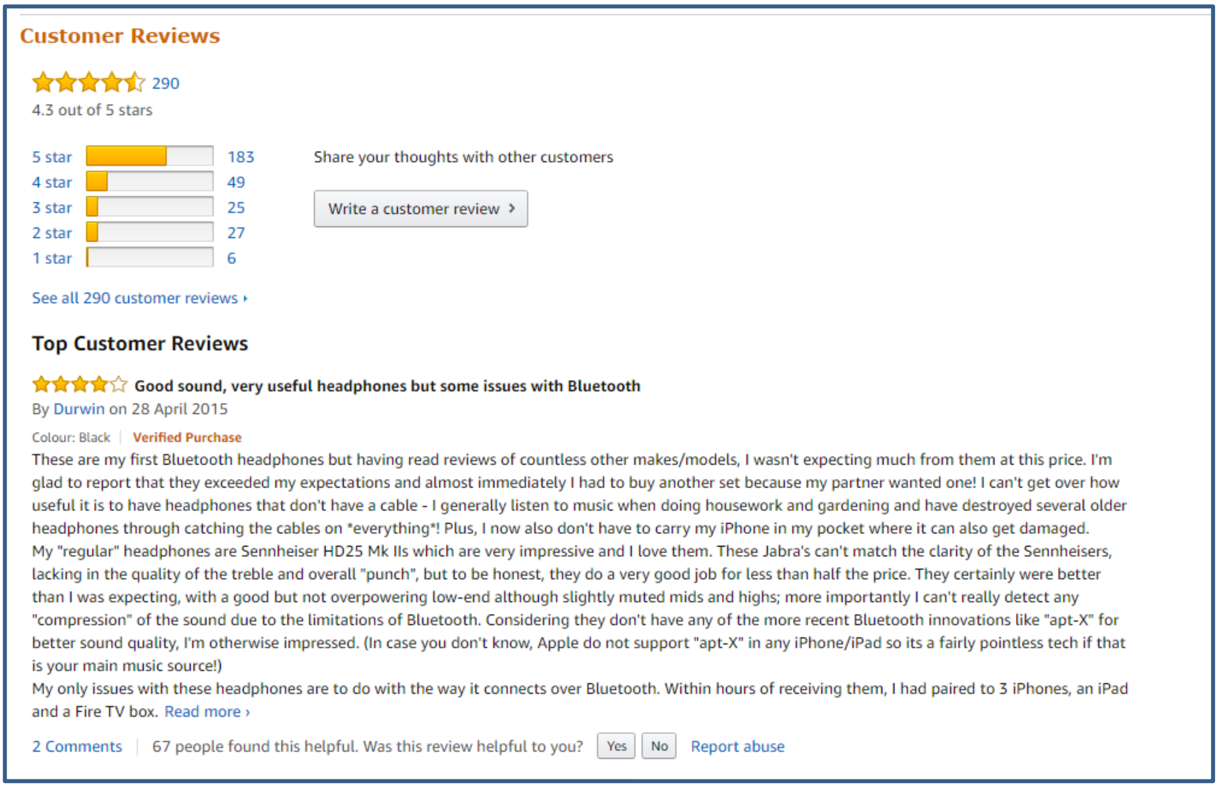


Figure 41 - Screenshot of product review from Amazon.com

Moreover, the fact that social platforms provide a sense of community to consumers, where they feel the responsibility to share information about products and help others place their purchase decision on that information. It also enables customers to interact with other customers on the platform such as rating the comments or reviews and finding the information helpful. This develops a sense of community between the current and potential customers and adds credibility to the user generated content on these platforms.

6.1.3 Perceived Quality

This research helps us understand the aspect of perceived quality and the features that were important to Jabra are whether the part of consumer's views. The initial list that was provided by Jabra indicate key performance attributes (refer to table 1) as sound, power, connectivity, functionality, design, comfortability, software, price, perception and accessories. Jabra also mapped list of words/attributes related to each of these performance feature.

However, result from our second experiment based on OLS regression, shows that consumer reviews keyed on different components that were correlated to these performance features. These insight help us gain understanding into the customer perception of quality and most valued products when it comes to Jabra products. The result shows that the performance indicator "price" in our experiment is highly correlated to "cheap", "expensive" and "cheaper". This provides Jabra with insights that consumer reviews might be talking about high price, or expensive products. However, as these words are missing the context, several perceptions can be derived from these correlated words based on the contextual meaning. The table mentioned below is the summary of OLS regression of key performance indicators along with top 03 (three) tokens appearing in high correlation based on coefficient value. The words (tokens), define customer reviews perception about the key performance indicators for example, reviews mention "loose", "causes" and "legend" in high correlation to feature "Comfortability". Although these tokens do not carry a lot of sentiment for naive researchers, we believe that it brings great insight for Jabra's product development, quality management and other units to analyze as what is perceived by these words in correlation to performance indicator features.

Feature	Tokens	Coefficients		Feature	Tokens	Coefficients
Accessories	charger	0.158		Software	happens	0.227
	waiting	0.154			directly	0.173
	turns	0.144			version	0.164
Comfortability	loose	0.167		Power	warning	0.375
	causes	0.146			turn	0.265
	legend	0.142			wont	0.224
Connectivity	turns	0.362		Perception	item	0.266
	died	0.342			return	0.231
	pair	0.328			amazon	0.222
Design	reset	0.181		Price	cheap	0.220
	hard	0.159			expensive	0.186
	hardware	0.142			cheaper	0.120
Functionality	poor	0.203		Durability	warranty	0.422
	chance	0.180			broken	0.379
	initially	0.137			defective	0.263
Sound	music	1.014		Others	heart	0.131
	sound	0.910			sensor	0.118
	allow	0.904			rate	0.112

Table 4 - Summary of top 3 tokens with highest coefficients per key_point

6.1.4 Customer satisfaction

As explained above in section of conceptual framework, customer satisfaction is related to post purchase behavior when customers have experienced the product and thereafter, share their experience. Jabra's quality management team is also interested in customer satisfaction. Customer satisfaction when measured through positive reviews plays a central role for organization's strategy for marketing and product improvement. It is a good predictor of purchase behavior that might lead to repurchase, purchase intentions and switching products behavior. Therefore, it is important to base quality management decisions based on results from BI tools dashboard presented in the table below.

Feature	Product ID with –ive polarity
Accessories	71, 85, 41, 23
Comfortability	41, 38, 37
Design	52, 9, 48, 64
Functionality	74
Sound	52
Connectivity	50, 80, 52, 42
Software	66, 77, 74, 85, 16, 60, 38, 45, 52
Power	41, 42, 39
Perception	66, 38, 45, 16, 37, 25, 8, 52, 60
Price	74
Durability	55, 19, 52, 63, 5, 9, 89, 25, 48, 66, 23
Others	41, 1

Table 5 - Summary of product IDs with negative polarity per key_point

We believe that assessing the product IDs with negative polarity (right column of the table 5) in regards to the features as performance indicators (left column of the table 5) carries more information as which products got most negative reviews considering the specific feature. Jabra can take advantage of this kind of information to analyze how the products can be improved in the context of these features. Thus, by understanding the products and features with highly negative polarity, Jabra can improve the quality management issues, resulting in customer satisfaction.

6.1.5 Automated market research

The purpose of the research is to apply automation techniques and text-mining algorithms to our dataset, in order to facilitate the analysis of market structure in two ways. First, to understand perception of consumer about Jabra products as presented in user-generated comments on Amazon.com and Bestbuy. Second, the mass production of opinions on the Internet is represented in the continuous stream of reviews. This provides practical grounds to shift from traditional approaches towards automated market research techniques that are continuous, automatic, inexpensive, and in real time.

Jabra classify data manually from the web-source, it also manually labeled the reviews into categories which is a lengthy and time consuming process. However, we believe that automated sentiment analysis would help Jabra to focus on important aspects or attributes related to performance indicators with the help of BI tools and dashboard to run data in real time, that is more cost effective and time efficient. Moreover, visualization features like “word cloud” can also help quality management team to focus on most important aspects and features of product performance.

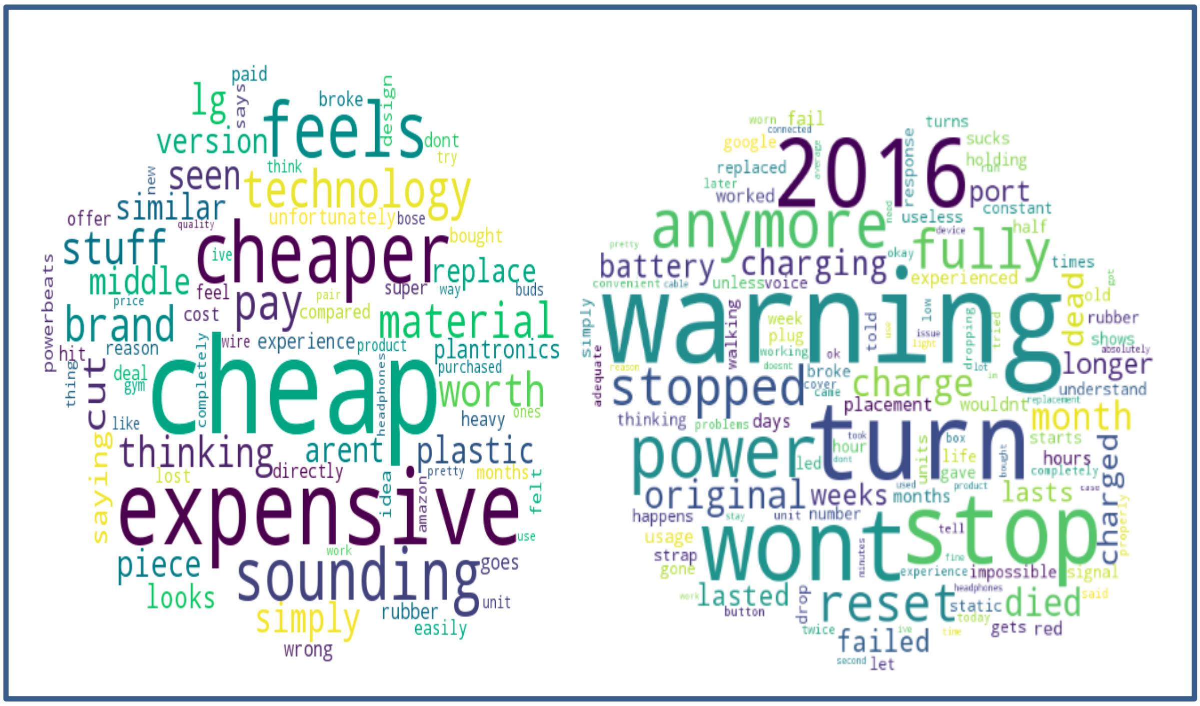


Figure 42 - Sample of word clouds

The above “word clouds” are simplest act of visualization of performance indicators based on customer reviews. For example, words that appear big in the cloud indicate higher correlation in regard to the whole dataset. Which infers that, consumer reviews when talking about “price” as performance indicator, they use words like “cheaper”, “cheap”, “expensive” and “sounding” in their reviews to share their experience. On the other hand, the “word cloud” for performance indicator “power”, highlights words as “warning”, “power”, “won’t stop”, “reset” and “died”. Several meanings could be inferred depending on different business units of Jabra. However, we are interested in quality management aspect for the sake of this research; therefore, we assume that Jabra’s quality management team assesses these word clouds and consequently would improve the product quality.

6.2 Answers to the Research Question(s)

By introducing an automated method to capture customer opinions and assess the product quality based on consumer reviews retrieved from e-commerce platforms, this research has provided a direct answer to the research question: "*What is the effect of text mining techniques on quality management of Jabra, using consumers' product reviews?*". We have successfully applied various text mining techniques in the data preprocessing and analysis stages, and with the support of QlikSense as a BI tool, we can provide the product assessment in a simple, straightforward and reliable manner for the case company. The process has been evaluated against the existing manual approach by the company, proving the effectiveness of our recommended method. Therefore, it can be inferred that text mining techniques would improve the quality management process for Jabra management.

Corresponding to the sub-research questions, this research has (i) developed a sentiment analysis approach which minimizes manual error and cost effective, (ii) built an automatic measurement of feature correlation with words. For the sub-research question "*What is the effect of automation techniques using text mining on evaluating consumer sentiment?*" a sentiment analysis module is developed in Python, with its subordinating libraries. The new sentence approach is introduced, which resolves the concern about multiple opinions within one review. The sentiment result is displayed in term of sentiment labels, polarity and subjectivity values. Then, they are combined with BI tool to address the consumer opinions about products, and detect products which needs management attention. Thus, we can assume that text mining techniques provide detailed approach towards evaluating consumer sentiments.

In term of the sub-research question "*What is the effect of automation techniques on the process of extracting key features from reviews?*" the OLS regression is chosen as the method for keyword extraction. This algorithm is one of the most common regression techniques, and it only requires basic statistical knowledge to evaluate the result. By applying OLS regression, this experiment provides analysts with list of keywords correlated with product attributes, which ease the process of manual feature mapping. The POS tagging phase is also proven to be beneficial to provide manager with detailed information regarding product quality assessment. Provided that, it can be inferred that automation techniques are well-suited for the process of key features extraction from consumer reviews.

The case study in this research has presented an automatic approach to handle consumer opinions from e-commerce platforms, and translate them into business values. It supports management level not only in understanding their customers, but also in evaluating their product quality based on consumer reviews. Results from this case study can be used as a baseline for similar research or experiment within retail industry domain.

6.3 Implications for Research

The study of consumer opinions retrieved from e-commerce platforms has carried several implications for research. Firstly, the development of Web 2.0 and social media platforms have transformed the buying behavior of consumers in recent days. It has, therefore, posed a new challenge for researchers and analysts to study their behaviors in these modern platforms, especially when the two way communication is presently dominating the Internet. Secondly, the study suggests further research into user generated content and automatic market research. The advancement of technology has brought consumers with more power on the websites, so they are more open to express their opinions. Therefore, companies should pay more attention to data content, as well as methods to perform automatic market research using this data. Thirdly, the result of this research has proven that consumer feedback does not merely serve for understanding the consumers, but they are also effective in knowing product experience. Therefore, organization may perform further research regarding product quality assessment using social data.

Additionally, in term of data analytics techniques using for the research, the usage of text mining techniques on social data also opens floors for further research. Unlike the traditional text streams, languages used on social platforms carry more modern expression, slangs or abbreviations. As a consequence, it is required to have more in-depth study about the effects of exceptional units of language on extracting the consumer sentiment values.

6.4 Implication for Practice

For this research, we aimed to address the quality assessment concern of case company based on consumer opinions. We have done this by introducing an automatic approach using text mining techniques and supplementary tools. Accordingly, the first major practical contribution of the present research is that it provides much information regarding the consumer satisfaction levels among those who are using Jabra products and expressing their opinions on e-commerce platforms. By evaluating the consumer satisfaction, Jabra product quality manager can have an initial assessment over products, and have an overview about how community is responding to their products.

A second important implication of our study derives from our findings on review language by consumers when writing a review. By analyzing keywords with high correlation to product attributes, our study points towards specific sets of words which consumers usually use in online reviews about Jabra products. Especially, when analyzing reviews with low rate, we can provide sets of words, which can reveal some aspects of products that need management attention. The case company can hence utilize the results in assessing their product quality in practice.

A third implication stems from our suggested approach for addressing the business concern. We have used multiple text mining techniques with some supporting analytic tools in order to automate the processes of sentiment analysis and sentence categorization, which are currently performed manually by Jabra. Afterwards, the suggested approach is evaluated against the manual approach to assess its performance. The result of this research hence can be used in practice to ease the process of addressing online consumer reviews, and support the case company in assessing their product quality management.

6.5 Recommendations for Case Company

The result of this study suggests various recommendations for the case company. First and foremost, it has confirmed that the data retrieved from social platforms carry helpful information regarding consumer opinions. However, the dataset can give more significant insights for analysts if they carry statistic values about the number of consumers who agrees with the reviews. In fact, it is an important piece of information because reviews with high rate of approval can represent the opinions of the majority, thereafter, company can focus more on the "found helpful" reviews. Additionally, if information about users and their activities are included in the dataset, analysts can also detect fake reviews and improve the accuracy of experiments.

Another recommendation for the case study is applying the suggested automatic approach in sentiment analysis and sentence categorization. The suggested approach has been assessed against the manual approach in term of its performance, so the case study can apply it in practice in order to minimize the manual error and costs.

Additionally, the case company could use a BI tool to perform data analytics with data visualization, and quickly data slice-and-dice for executive review. Also, it is a supplementary tool in assessing the product quality. For instance, the data analytics on QlikSense using sentiment results can reveal that consumers do not hold any negative feedback for product with id 7, except for the price (figure 43). Thus, by combining results from text mining techniques with BI tool, the case company can reveal different aspects of the consumer opinions.

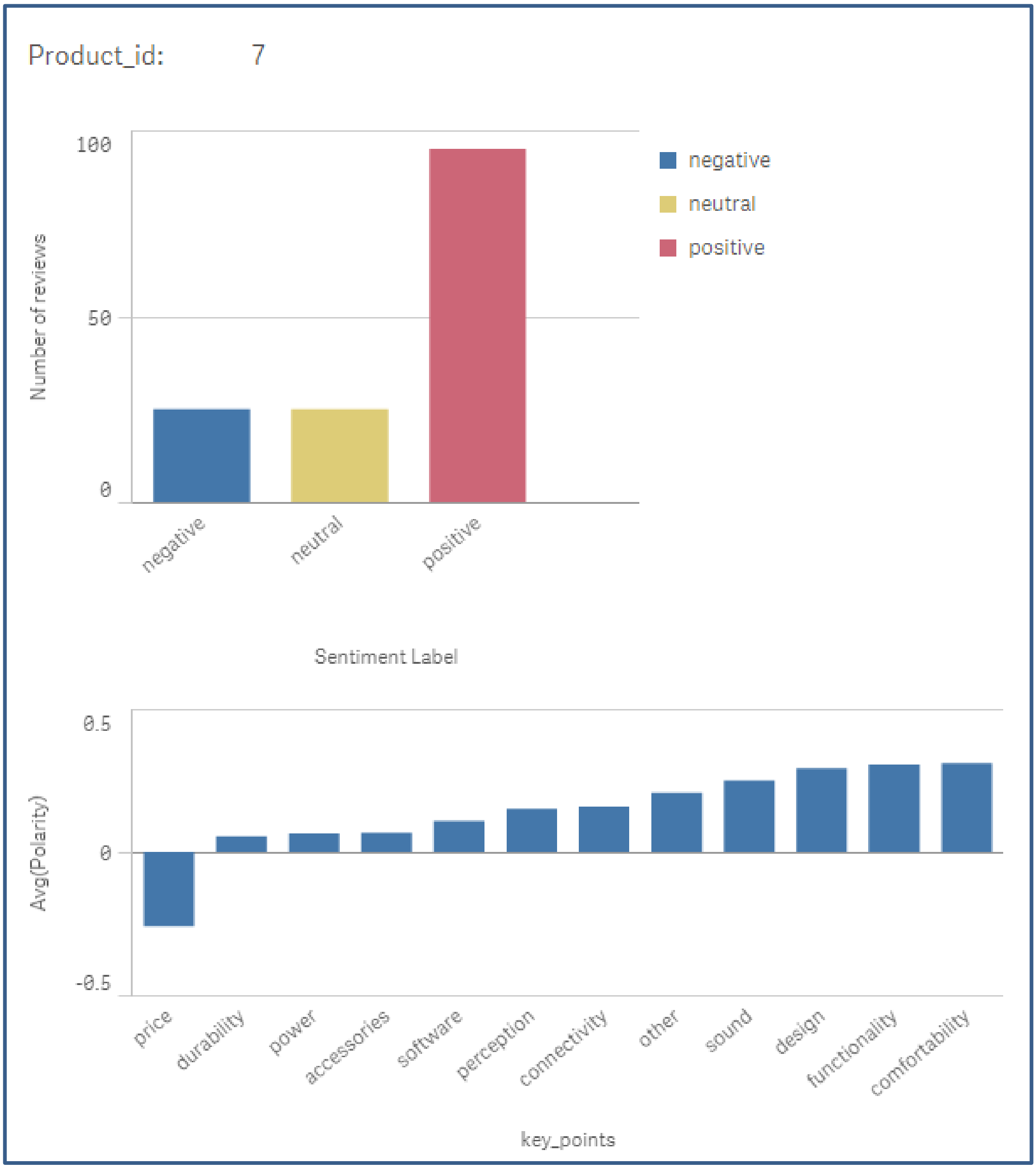


Figure 43 - Dashboard of sentiment result for product id 7

6.6 Limitations of the Study

The presented results in this project needs to be considered in the context of limitations. Also, the process of addressing and answering our research questions basically generates some more aspects that can be explored in further research. This paper presents few limitations in relation to our experiments and we would like to discuss these limitations in brief summary.

Consumer reviews based on social media platforms are written in informal language that often contains slangs and abbreviations. To certain extent, slangs and abbreviations could be processed by analytical tools. However, when it comes to labeling or classifying data, there might be issues correctly classifying them due the fact that the NLTK language corpus is based on standard English words, while slangs and abbreviations are results of modern communication on social media platforms e.g. lol, omg, rofl, btw etc.

Another limitation of our research was sarcasm detection, as some of the customer reviews might carry some sarcasm in the tone, which could be understood within the context of review. However, once we split the reviews into sentences, it is difficult for the NLTK library to detect the sarcasm in regard to the review.

For the purpose of this research, the data was derived from Amazon.com and Bestbuy. We believe that one of the limitations for the study could be two similar data sources. Data platforms such as, Facebook, blogs and Jabra's official website could have brought new insight into our experiments, as different platforms offer different feature affordances to facilitate users to generate reviews. For example, Facebook has the ability to like, comment and share the reviews if found helpful.

Furthermore, due to the time constraint, we did not build up our own sentiment classifier. Instead, we could only run the sentiment analysis on Textblob. This has impacts on our results, since we performed analysis based mostly on the results of sentiment analysis.

Also, in the regression analysis experiment, we only took unigram into consideration. Single words are used as tokens to examine the relationship between words and the topics. However, it will not provide as much information as n-grams (with n >1) technique could have helped. For instance, if we perform analysis on unigram, the word "cheap" may refer to the price of product, but the bigram "cheap quality" may imply the poor quality of product. Therefore, n-grams like bigrams or trigrams may yield significant improvement in the research.

Another limitation of this research is that there was missing element of credibility on the reviews i.e. how many people find the review helpful. Although, Amazon.com has feature of "found this review helpful", our data set did not include this feature. We believe that if we could add the aspect of “found this review helpful”, it would have added to the credibility of the reviews.

Last but not the least, a limitation in our dataset could be fake reviews. In order to achieve accuracy in our results, it is critical to detect and eliminate irrelevant or fake reviews. However, there is no supporting data to perform the data filtration.

6.7 Future Work

This section presents possible future work that includes natural language processing techniques and possibilities for the use of NLTK as a secondary tool for opinion mining, as well as further steps which could be taken to use tools like MUTATO and SentiWordNet as primary tools for sentiment analysis.

Slangs and abbreviations:

Slang is a type of expression which is very informal and used mostly in chats and user generated content on social media platform. For example, “lol” is a slang which is commonly used to express the emotion of “laughing out loud”. Abbreviations are shortened form of words and phrases which are used in common language to save time. For further research, data mining techniques can be applied to explicitly address slangs and abbreviations in order to retrieve more information from the reviews.

Sarcasm Detection:

Sarcasm is a way to mimic or convey dislike. Opinions on social media platforms are used for “opinion mining” and the use of sarcasm could mislead readers by giving different or opposite views than the understood meaning. For further research we need to develop a model to train the dataset based on sarcasm detection.

Spelling Mistakes:

Spelling mistakes are quite common in user-generated content as these are written informally without strong focus on spelling correction. Often spelling mistakes are made by skipping some characters from word or interchanging the alphabets order. For further research we are interested in addressing the problem of spelling in order to maximize the output of text mining.

N-gram analysis:

Regarding the text data analytics, we can extend the experiment by analyzing the n-grams (such as bigrams or trigrams) instead of unigrams. It is helpful to include the context awareness in the research, as it can provide more precise information about consumer opinions.

Other tools and data sources:

Further studies could be performed in order to explore the result when running data from other data sources such as Facebook, Twitter and blogs. It would also be interesting to explore the reviews from the Jabra official website. This will provide us with results that can be compared in order to measure efficiency of tools and techniques. Moreover, we can use tools like MUTATO and SentiWordNet to do the analysis and compare results with NLTK.

Detecting fake reviews:

Further research could be done on fake reviews as it is also an interesting aspect to spot fake reviews from the data set. This will improve the data accuracy and spot actual reviews with complaints along the rating and detect fake reviews with the help of review length, tone and emotional language and frequency of reviewer to submit reviews.

CHAPTER 7

Conclusion & Reflection

7. Conclusion & Reflection

In this chapter, we would like to presents the conclusion of our research and significant learning reflections along the process of conducting this research.

7.1 Conclusion

Our research aimed at analyzing consumer opinions about Jabra product quality management aspects using the online consumer reviews that are usually posted and discussed on e-commerce websites i.e. Amazon.com and BestBuy. Jabra along with most retail stores encourage their consumers to write products reviews on their websites. However, consumer communication is no longer limited to one medium for user reviews and rating, when a large number of consumer reviews can be retrieved from common e-commerce websites. This shift has brought companies involving in the process to expose to a new challenge of understanding their consumer expectations, which recently vary on different platforms.

Our aim is to develop valuable insight for Jabra to take advantage from the pool of unstructured information, that when analyzed properly can contributes to the transformation of product development and create customer satisfaction. By having deep insights into various consumer experiences, the organizations can have an exact evaluation of their product quality, and perhaps design better system functionalities in future product release.

Based on explained above insights regarding consumer reviews and ease of use related to social platforms, this study focuses on the aspects of user generated content and product quality assessment using text mining techniques. The dataset of consumer reviews is used for two primary experiments. Firstly, we perform sentiment analysis about customer’s comments on social platforms in order to understand how consumers express their opinions about Jabra products. Secondly, we evaluate specific features with their coefficient value of tokens, and assess the quality of features as performance indicators.

Finally, we evaluate the performance of the suggested approach by understanding the results in light of the manual approach by the case company. It is important to acknowledge that the focus of the study is not comparing the manual approach to our automatic text mining techniques, but we used the manually labeled data as benchmark to measure the efficiency of our results.

The research aims to analyze and address problems regarding product quality management and deliver significant outcomes based on data provided by Jabra. Therefore, it was a milestone to present to the case company in order to convey our findings from experiments. The results were presented to the Jabra quality management team at the workshop held in Copenhagen Business School campus[2] in order to discuss the outcomes and granularity of the analysis.

During the workshop with Jabra, the quality management team found the sentences level approach useful as to point out towards specific attributes. Moreover, Jabra team is interested in using the results from OLS regression to study words as tokens to improve their list of key performance indicators. The organization is also interested in deploying the text mining techniques in order to take advantage of customer reviews by running algorithms on real time data. This will help them run the similar techniques on vast amount of data, improving the experiment efficiency.

7.2 Learning Reflections

The research has helped us gain insight that the process of sentiment analysis is definitely a challenging area. Although we have utilized many powerful tools like NLTK, Textblob, sklearn and stats models for this research, we still believe that the processing is insufficient to capture the real meaning of the terms. Human language is often ambiguous, thus, it may need new algorithm or methods to process them. Especially, the language used on the social media platforms is informal with slangs and abbreviations and is continuously growing due to increased use of ubiquitous computing devices; it causes more difficulties for analysts to handle it.

Another interesting insight is that, consumer reviews on social network platforms tend to be longer when they carry negative comments than positive ones. It could be inferred that dissatisfied customers write longer reviews including several aspects

[2] 5th of May 2017

and explanations about product features and experiences. Organizations can take advantage of these reviews because there is room for improvement based on these complaints.

References

Baruah, T. D. (2012). Effectiveness of Social Media as a tool of communication and its potential for technology enabled connections: A micro-level study. *International Journal of Scientific and Research Publications, 2(5), 1-10.*

Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python---Analyzing Text with the Natural Language Toolkit O'Reilly Media.

Cambria, E., Havasi, C., & Hussain, A. (2012, May). *SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis*. In FLAIRS conference (pp. 202-207).

Chen, C. H. (2016). Improved TFIDF in big news retrieval: An empirical study. *Pattern Recognition Letters*.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, *36*(4), 1165-1188.

Chen, T., Tribbitt, M. A., Yang, Y., & Li, X. (2017). Does rivals' innovation matter? A competitive dynamics perspective on firms' product strategy. *Journal of Business Research*, *76*, 1-7.

Chen, P., S. Wu. 2004. The impact of online recommendations and consumer feedback on sales. Proc. 24th Internat. Conf. Inform. Systems, Washington, D.C.

Chevalier, J. A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," Journal of Marketing Research (43), pp. 345-354.

Cullen, J. (2007). Information work and the opportunity of innovation From corporate to social product development. *Business information review*, *24*(3), 156-160.

Danninger, M., Vertegaal, R., Siewiorek, D. P., & Mamuji, A. (2005, May). Using social geometry to manage interruptions and co-worker attention in office environments. In *Proceedings of Graphics Interface 2005* (pp. 211-218). Canadian Human-Computer Communications Society.

Das, S. and Chen, M., 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. APFA'01.

Dellarocas, C. 2006. Strategic manipulation of Internet opinion forums: Implications for consumers and firms. Management Sci. 52(10) 1577–1593.

Dismuke, C., & Lindrooth, R. (2006). Ordinary least squares. *Methods and Designs for Outcomes Research*, *93*, 93-104.

DoubleClick. 2004. DoubleClick's touchpoints II: The changing purchase process.

Filieri, R. (2015). What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research*, *68*(6), 1261-1270.

Golfarelli, M., Rizzi, S., & Cella, I. (2004, November). Beyond data warehousing: what's next in business intelligence?. In *Proceedings of the 7th ACM international workshop on Data warehousing and OLAP* (pp. 1-6). ACM.

Godes, D. and D. Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," Marketing Science, 23 (4), 545–60.

Hallowell, R. (1996). The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study. *International journal of service industry management*, *7*(4), 27-42.

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014, January). Word cloud explorer: Text analytics based on word clouds. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 1833-1842). IEEE.

Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.

Hussein, D. M. E. D. M. (2016). A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*.

Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Sage.

Janyce M Wiebe, Rebecca F Bruce, and Thomas P O'Hara. Development and use of a gold-standard data set for subjectivity classifications. In Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, pages 246–253. Association for Computational Linguistics, 1999.

Jin, W., Ho, H. H., & Srihari, R. K. (2009, June). OpinionMiner: a novel machine learning system for web opinion mining and extraction. In *Proceedings of the 15th ACM SIGKDD* *international conference on Knowledge discovery and data mining* (pp. 1195-1204). ACM.

Jumin Lee, Do-Hyung Park, Ingoo Han, (2011) "The different effects of online consumer reviews on consumers' purchase intentions depending on trust in online shopping malls: An advertising perspective", Internet Research, Vol. 21 Iss: 2, pp.187 - 206

Kaplan, B., & Duchon, D. (1988). Combining Qualitative and Quantitative Methods in Information Systems Research: A Case Study. *MIS Quarterly, 12*(4), 571-586. doi:10.2307/249133.

Krippendorff, K. Content analysis: an introduction to its methodology. Sage, 2004.

Kumar, N., & Benbasat, I. (2006). Research note: the influence of recommendations and consumer reviews on evaluations of websites. *Information Systems Research*, *17*(4), 425-439.

L Dini and G Mazzini. Opinion classification Through information extraction. Intl. Conf. on Data Mining Methods and Databases for, 2002.

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, *48*(5), 881-894.

Lee, T. H., Collmeyer, A. J., & Wong, D. T. (2009). *U.S. Patent No. 7,548,040*. Washington, DC: U.S. Patent and Trademark Office.

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, *19*(4), 456-474.

Lin J. Design Optimization of Headband for Headphone. Dissertation. Washington State University. 2009.

Liu Bing Hu Minqing. Mining and summarizing customer reviews. KDD'04, 2004. vii, 27, 30, 31, 32, 33, 34, 35, 38, 59.

Liu, S., Zhou, M. X., Pan, S., Song, Y., Qian, W., Cai, W., & Lian, X. (2012). Tiara: Interactive, topic-based visual text summarization and analysis. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *3*(2), 25.

Malthouse, E. C., Haenlein, M., Skiera, B., Wege, E., & Zhang, M. (2013). Managing customer relationships in the social media era: introducing the social CRM house. *Journal of Interactive Marketing*, *27*(4), 270-280.

Michael Milton. Head First Data Analysis: A Learner's Guide to Big Numbers, Statistics, and Good Decisions. O'Reilly Media, 1 edition, 2009.

Mitkov, R. (2005). *The Oxford handbook of computational linguistics*. Oxford University Press.

Moutinho, L., & Hutcheson, G. D. (Eds.). (2011). *The SAGE dictionary of quantitative management research*. Sage.

Mozer, T. F., & Mozer, F. S. (2012). *U.S. Patent No. 8,195,467*. Washington, DC: U.S. Patent and Trademark Office.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon. com.

Negash, S., & Gray, P. (2008). Business intelligence. *Handbook on decision support systems 2*, 175-193.

Olshavsky, Richard W. (1985), "Perceived Quality in Consumer Decision Making: An Integrated Theoretical Perspective," in Perceived Quality, J. Jacoby and J. Olson, eds. Lexington, MA: Lexington Books, 3-29.

Pang, B., & Lee, L. (2004, July). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *In Proceedings of the 42nd annual meeting on Association for Computational Linguistics (p. 271)*. Association for Computational Linguistic

Park, S. H., Huh, S. Y., Oh, W., & Han, S. P. (2012). A social network-based inference model for validating customer profile data. *MIS Quarterly*, *36*(4), 1217-1237.

Pavlou, Paul A. and David Gefen (2004), "Building Effective Online Marketplaces with Institution-Based Trust," Information Systems Research, 15 (1), 37–59.

Perkins, J. (2014). *Python 3 Text Processing with NLTK 3 Cookbook*. Packt Publishing Ltd.

Piller, F. T., Vossen, A., & Ihl, C. (2011). From social media to social product development: the impact of social media on co-creation of innovation.

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. " O'Reilly Media, Inc.".

Resnick, P. and R. Zeckhauser (2002), "Trust Among Strangers in Internet Transactions:

Empirical Analysis of eBay's Reputation System," in Advances in Applied Microeconomics: The Economics of the Internet and E-Commerce, Vol. 11, Michael R. Baye, ed. Amsterdam: Elsevier Science, 127–57.

Richins, M. L. 1983. Negative word-of-mouth by dissatisfied consumers: A pilot study. J. Marketing 47 68–78.

Samaneh Moghaddam and Martin Ester. Aspect-based opinion mining from online reviews. In Tutorial at SIGIR Conference, 2012.

Satoshi Morinaga, Kenji Yamanishi, Kenji Tateishi, and Toshikazu Fukushima. Mining product reputations on the Web. International Conference on Knowledge Discovery and Data Mining, page 341, 2002.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, *8*(9), e73791.

Senecal, Sylvain and Jacques Nantel (2004), "The Influence of Online Product Recommendations on Consumers' Online Choices," Journal of Retailing, 80 (2), 159–69.

Stake, R. E. (1995). *The art of case study research*. Sage.

Trompette, N., & Chatillon, J. (2012). Survey of noise exposure and background noise in call centers using headphones. *Journal of occupational and environmental hygiene*, *9*(6), 381-386.

Wang, J., Zhao, J., Guo, S., North, C., & Ramakrishnan, N. (2014, May). ReCloud: semantics-based word cloud visualization of user reviews. In *Proceedings of Graphics Interface 2014* (pp. 151-158). Canadian Information Processing Society.

Wang, T. (2017). Social identity dimensions and consumer behavior in social media. *Asia Pacific Management Review*.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347-354). Association for Computational Linguistics.

Yergeau, F. (1996). *UTF-8, a transformation format of Unicode and ISO 10646* (No. RFC 2044).

Zeng, D., Chen, H., Lusch, R., & Li, S. H. (2010). Social media analytics and intelligence. IEEE Intelligent Systems, 25(6), 13-16.

Zeithaml, V.A. (1988) Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. Journal of Marketing, 52, 2–22

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing*, *74*(2), 133-148.

Links

http://www.jabra.com/about

https://en.wikipedia.org/wiki/Jabra_(headset)#cite_ref-1

Appendices

Appendix A – Original dataset

Appendix B – List of keywords classified by Jabra

Appendix C – Preprocessed dataset

Appendix D – Sentiment analysis result

Appendix E – OLS regression analysis result

Please refer to the following link for the appendices:

https://drive.google.com/open?id=0B_pDhbTB529WQk5OT3drT3pEb1E