UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI ECONOMIA "GIORGIO FUÀ"

Corso di Laurea Magistrale in
International Economics and Commerce
Curriculum Business Organization and Strategy

ONLINE CONSUMER REVIEWS IN E-COMMERCE: THE ROLE AND IMPACT ON BUSINESS PERFORMANCE

LE RECENSIONI ONLINE DI UN SITO DI E-COMMERCE: A COSA SERVONO E COME IMPATTANO LA PERFORMANCE AZIENDALE

Relatore: Chiar.ma
Prof.ssa Sara Bartoloni

Tesi di Laurea di:
Laura Campetella

Correlatore: Chiar.ma
Prof.ssa Claudia Pigini

Anno Accademico 2020 – 2021

Table of Contents

ABSTRACT

Il tema della tesi è quello di indagare il ruolo delle recensioni di un sito di e-commerce e il loro impatto sulla performance aziendale. Il passaparola è la miglior arma per diffondere informazione, ed è per questo che le recensioni online stanno acquistando sempre più importanza. È stato dimostrato che i consumatori, prima di effettuare un acquisto, si affidano molto alle review scritte da altri utenti che hanno acquistato in precedenza il prodotto o servizio d'interesse.

Oggigiorno, le recensioni sono diventate un pilastro del marketing ed è fondamentale analizzarle per comprendere al meglio il comportamento dei consumatori e l'influenza che hanno su di essi. Le review sono anche uno strumento utile per costruire la "*brand reputation*", in quanto sono lo specchio di come i consumatori percepiscono l'azienda e i suoi prodotti o servizi. Il successo, la reputazione e la credibilità sono infatti direttamente proporzionali ai feedback dei suoi consumatori: le aziende che possiedono un e-commerce tendono sempre a mostrare al proprio pubblico digitale le recensioni ottenute, così da dare nell'immediato un riscontro reale sulla propria credibilità. Un numero elevato di recensioni positive si può tradurre in un aumento delle vendite, così come una quantità elevata di recensioni negative possono causare una riduzione delle vendite.

Lo studio presenta anche un esempio pratico di come i proprietari di un e-commerce o di un sito web possono applicare la text e la sentiment analysis per analizzare le recensioni dei propri consumatori. Grazie all'analisi, sono emerse molte informazioni utili relative ai punti di forza e debolezza dell'azienda che permettono di individuare gli aspetti da potenziare al fine di soddisfare al meglio i bisogni dei consumatori.

ABSTRACT

The aim of the thesis is to investigate the role of the online consumer reviews of an e-commerce website and their impact on business performance. Word-of-Mouth is the most effective mean to spread information, hence, online reviews are now more important than ever. It has been demonstrated that consumers, before making a purchase, usually rely on feedback written by individuals who have already purchased the product or service.

Reviews have become a foundation of marketing in recent years, and it is critical to analyze them to better understand customer behavior and the impact they have on consumers. Recommendations are also an effective technique for establishing *brand reputation*, since they reflect how customers perceive a firm and its products or services. The success, reputation, and credibility of a company are directly proportionate to the reviews it receives from its customers: e-commerce owners frequently show their digital audience the reviews they have received, in order to provide rapid feedback on their trustworthiness. A large number of positive reviews can lead to an increase in sales, as well as a great number of negative reviews can cause a sales reduction.

The study also includes a demonstration of how e-commerce businesses might use text and sentiment analysis to analyze their customer evaluations. A lot of important information about the company's strengths and weaknesses has emerged as a result

of the analysis, allowing the firm to identify the aspects that need to be enhanced in order to best satisfy its customers' needs.

INTRODUCTION

The aim of this thesis is to demonstrate that e-commerce owners can use online consumer reviews to get valuable insight into customers' preferences and sentiments over the products and services provided by their company, as well as their overall performance. Indeed, this new form of communication, also known as Electronic Word-of-Mouth, is perceived by consumers as being more credible than the information supplied by companies, since it is conveyed by individuals who have actually purchased the item or service.

Furthermore, data analysis strongly affects companies' decisions. Firms can track the customer engagement, target users based on purchase history, and even forecast consumer demand to run more efficient campaigns, as well as identify the main issues from the complaints made by customers to solve them and better satisfy the consumers' needs.

Chapter 1 will present an overview of the concepts of digital marketing and big data, providing an explanation of how the two subjects are related and how they influence the different touchpoints of the customer journey. Indeed, customer satisfaction is one of the most important factors in customer retention, thus successful businesses measure it on a regular basis. Furthermore, by defining important assessment measures, data allows firms to identify whether marketing operations are doing well or not, allowing managers to implement the most successful marketing strategy.

Chapter 2 will provide a definition of online consumer review and word-of-mouth and will investigate the relationship between reviews and consumers' purchasing behavior to examine how they affect a company's sales. Indeed, this is an issue of critical relevance to businesses that sell their products or services online. Word-of-mouth communication is an important facilitator of learning and can have a huge impact on consumer's purchasing decisions. Many studies indicate the influence that this electronic source of information may have on the sales of the product or service reviewed.

To prove the importance of online consumer reviews and the insights a company can gain from data analysis, Chapter 3 presents a practical approach of text analytics and sentiment analysis conducted with RStudio. The intention is to provide an example of how reviews can be analyzed according to the information a company might want to understand. The dataset chosen to undertake the analysis presents the reviews of an e-commerce clothing platform extracted from Kaggle, a data science and machine learning online platform where users can search and post datasets, collaborate with data scientists and machine learning specialists, and compete in online challenges.

CHAPTER 1 - DIGITAL MARKETING AND BIG DATA FOR E-COMMERCE

1.1. DIGITAL MARKETING

In today's competitive world, marketing strategies play such an important role that may depict the success or failure of companies. Thanks to the evolution of the Internet, web, and other digital technologies, marketing is constantly changing. Dave Chaffey and Fiona Ellis-Chadwick described Digital Marketing as "the application of the Internet and related digital technologies in conjunction with traditional communications to achieve marketing objectives" (Chaffey and Ellis-Chadwick, 2016, p. 11).

Building a solid online presence is critical for an organization's sustainability. Marketers must first determine which digital innovations are most significant for the company, so that they can integrate them successfully with conventional marketing communications and get a competitive advantage.

Digital marketing is about managing multiple forms of online presence, such as websites and social media pages, together with new communication approaches, which assist companies in gaining new customers, building consumer preferences, promoting brands, retaining customers, and increasing sales. Such new communication techniques are, for instance, Social Media Marketing, SEM (Search Engine Marketing), online advertising, and DEM (Direct Email Marketing), and they contribute to the development of customer relationships through E-CRM (Electronic Customer Relationship Management).

Digital marketing does not completely differ from traditional marketing: the main critical choices are similar. However, digital marketing permits new kinds of engagement and alternative models for information transmission that strongly differ from traditional marketing communications. The main differences between digital and traditional marketing may be summarized in the "6 Is of the e-marketing mix" (McDonald and Wilson, 1999). The six Is are:

- Interactivity: through traditional media firms deliver a message to consumers, therefore they are considered push media, while online media use a pull mechanism, because customers are more likely to start contact with the firm and seek information by visiting the website.
- Intelligence: the Internet may be used to collect marketing research, particularly regarding customer views of products and services, at a minimal cost, so it produces two-way feedback that is seldom seen in traditional forms of communication.
- Individualization: interactive marketing communications may be tailored to the individual, as opposed to traditional media, which broadcasts the same

message to everyone. The process of delivering individualized content via email or web pages is called *Personalization* and is based on the information gathered about website visitors.

- Integration: when evaluating the performance of a website, the role of the Internet in connecting with consumers and other partners can be examined from two angles: organization-to-customer and customer-to-organization direction. The former refers to Outbound Internet-based communications, that is the way in which companies send personalized messaging via website and email marketing to generate new leads and retain the existing customers, while the latter relates to Inbound Internet-based communications, where customers ask for information through web-based forms and e-mail.
- Industry restructuring: while planning an e-marketing strategy, companies should consider the role of intermediaries between them and their customers. Indeed, there are two processes called *Disintermediation* and *Reintermediation* which refer to the removal and introduction of such intermediaries that are, for instance, distributors and brokers.
- Independence of location: the Internet allows companies to sell and communicate with consumers into overseas markets where they don't have sales or customer service personnel.

1.2. A NEW PATH TO PURCHASE: THE CUSTOMER JOURNEY

A purchase does not necessarily mean the end of the search for a customer. Due to the fact that consumers today connect with businesses through a multitude of touch points across numerous channels and media, research continues long after a purchase has been made, resulting in more sophisticated customer journeys (Lemon and Verhoef, 2016).

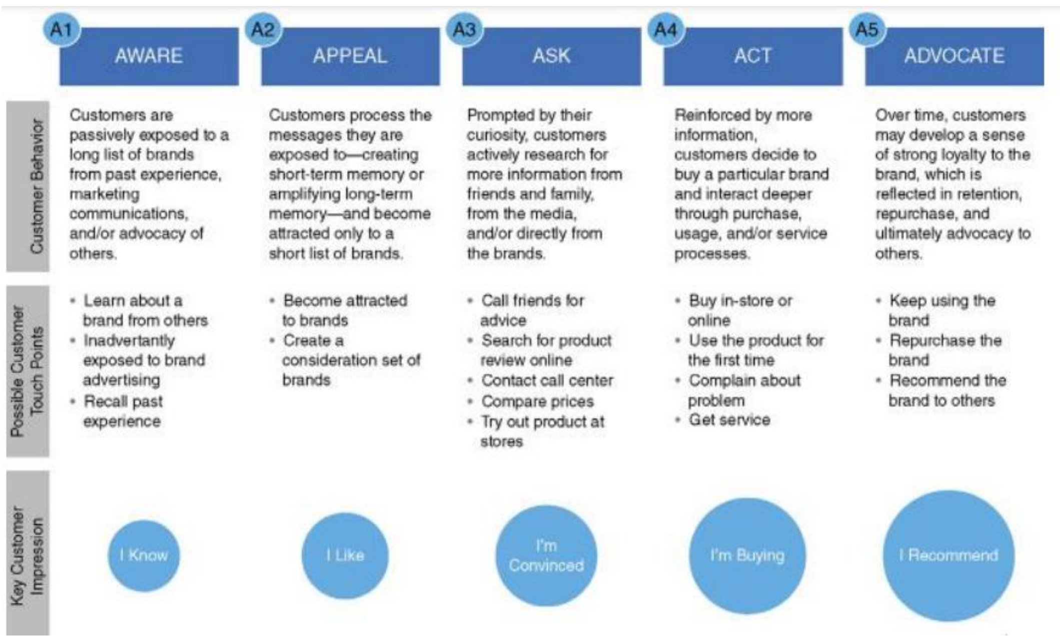
This sort of itinerary that the consumer follows when establishing a relationship with a company over time and in the different touchpoints, both online and offline, is called *customer journey*. A *touchpoint* can be defined as any individual contact between the company and the customer at distinct points in the buying experience. Customer experience, which is the internal and subjective response customers have to any direct or indirect contact with a company, encompasses all aspects of a company's offering, including the quality of customer care but also advertising, packaging, product and service features, ease of use, and reliability (Schwager and Meyer, 2007).

However, companies need to understand that having more touchpoints and higher volume in messages do not imply an increased influence (Kotler et al., 2017). As the economist Lawrence Abbot observed in 1955, "what people really desire are not products but satisfying experiences". Indeed, the best thing companies can do is to create a meaningful connection with their customers in a few key touchpoints, because it takes just one unusual positive moment to turn a customer into a loyal advocate.

The consumer path to purchase became much more complex due to digitalization and technology evolution, therefore, Kotler outlined the customer path with the so-called five A's: aware, appeal, ask, act, and advocate (Figure 1.1).

The *Aware* phase refers to the passive exposure of consumers to a huge number of brands. After this step, the *Appeal* phase starts, where customers process all the signals they have been exposed to and express interest for just a few brands. Customers enter the *Ask* phase when they become so curious that they start actively searching for information about the brands they are interested in. To learn more about a brand, they can also read online product reviews and try out products at brick-and-mortar stores. The convergence of the digital and physical worlds has made this phase considerably more challenging today, and since customers may seek information from various sources, companies must be present at least in the most popular channels (Kotler et al., 2017). This is a particularly critical phase, because customers will make decisions depending on what they learn through their interactions with others, that is why the brand attraction needs to be confirmed by other customers. If the information they get is compelling, consumers will choose to *Act*, which is not limited to the buying process, but also involves the post-purchase services. Therefore, companies must ensure that the ownership and usage experience is pleasant and unique. If customers build a strong feeling of loyalty to a brand, without being asked, they start actively advocating to promote the brand they enjoy, entering the stage of the *Advocate*, and they are also more inclined to purchase additional products or services from such companies.

Figure 1.1 – Kotler's Five A's



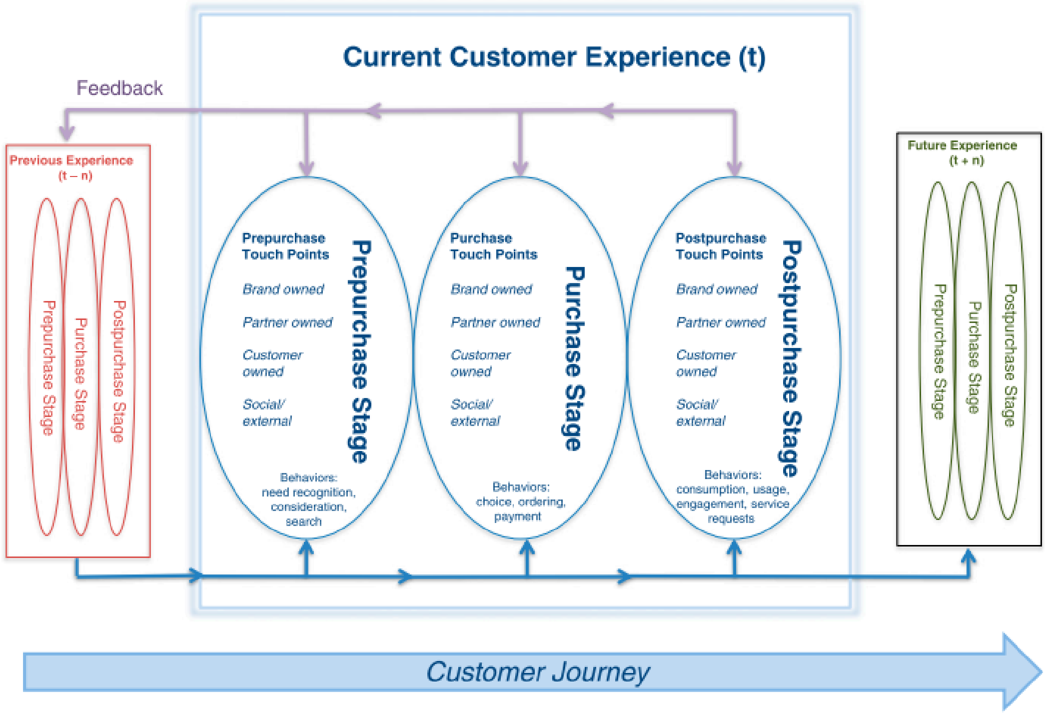
SOURCE: Kotler et al., 2017, p.64

Lemon and Verhoef conceptualize the total customer experience as a dynamic process, where the process flows from pre-purchase (including search) to purchase to post-purchase (Lemon and Verhoef, 2016) (Figure 1.2). The prepurchase stage refers to all the aspects of a customer's interaction with a brand prior to a purchase. Practically, however, this stage embraces the customer's experience from the initial awareness of a need/goal/impulse to the possibility of satisfying that need/goal/impulse with a purchase (e.g., Hoyer 1984; Pieters, Baumgartner, and Allen 1995; Lemon and Verhoef, 2016). All the interactions between the customer and the firm during the actual purchase activity, such as choice, ordering, and payment, are included in the purchase stage.

Finally, following the actual purchase, the post-purchase phase begins. It includes all the activities that are directly related to the brand or product/service, such as usage, consumption, and service request.

However, only some of the touchpoints in the customer experience are controlled by the firm. According to Lemon and Verhoef, there are "four categories of customer experience touch points: brand-owned, partner-owned, customer-owned, and social/external/independent" (Lemon and Verhoef, 2016, p.76), and the impact of each touch point class may vary according to the type of product/service or the customer's individual journey. Figure 1.2 provides a customer journey model designed by Lemon and Verhoef, which emphasizes the importance of the different touchpoints. Indeed, depending on where the customer is in the customer journey, each touchpoint takes on a distinct meaning (Pascucci et al., 2019).

Figure 1.2 - Process Model for Customer Journey and Experience



SOURCE: Lemon and Verhoef, 2016, p.77

Any brand-owned media (such as advertisements, social media, websites, and loyalty programs) as well as any brand-controlled marketing mix elements (from packaging to after-sales service) are of course included in the category of brand-owned touchpoints. Sometimes, firms create and manage touchpoints jointly with their partners, that may be marketing agencies, multichannel distribution partners and communication channel partners: this category is known as partner-owned touchpoints. On the contrary, the customer-owned touchpoints are just consumer behaviors that cannot be influenced or controlled by the company, its partners, or others. For instance, when customers evaluate their own needs and wants in the pre-purchase phase it is part of the customer-owned touchpoints, though partners may also play a role. However, they are more frequent and crucial in the post-purchase phase, when individual consumption takes center place. Nevertheless, social/external touchpoints are probably the most effective ones. Indeed, throughout the three stages of the experience, other people might exert a great influence on customers and may have a strong impact in the process, particularly during the buying phase.

Marketers have always been focusing on three critical moments that represent the mental model of consumers:

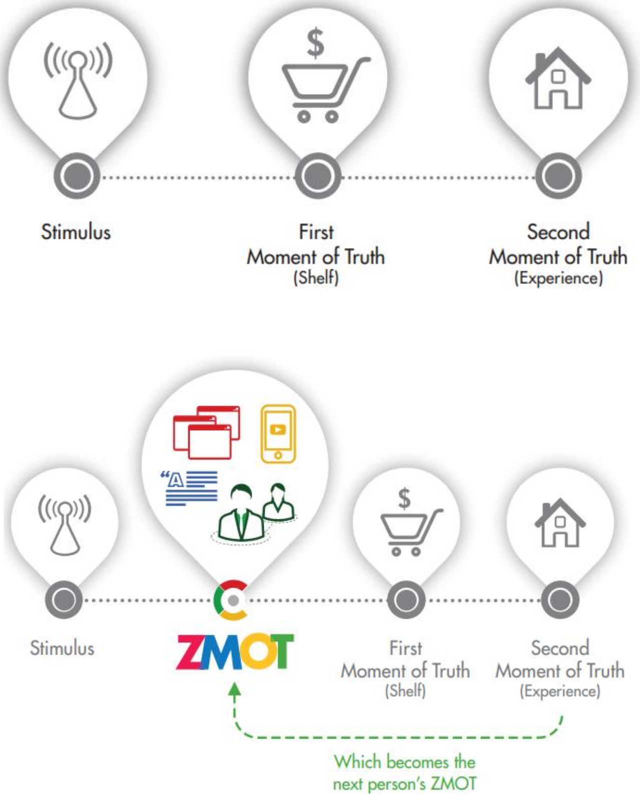
- *stimulus*, which arises when the consumer sees an advertising;
- *shelf*, that happens when the customer goes to the shop and buys the product seen on the ad;
- *experience*, when they use the product.

However, today consumers undertake a multichannel search before buying a product.

Google defined a New Mental Model, the so-called Zero Moment of Truth (ZMOT), which is that instant when consumers open their laptops, pick up their smartphones or grab their tablets, and search to see if you meet their needs

(Lecinski, 2011). In general, they get information by reading online reviews, blogs or social network discussions on the brand or product, or they can even watch some videos. The use of search engines and social media means that there are more options, so consumers will consider a wider range of items (Chaffey and Ellis-Chadwick, 2016). Figure 1.3 demonstrates the difference between the traditional and the new Mental Models developed by Google.

Figure 1.3 - The Traditional 3-Step Mental Model vs. the New Mental Model



SOURCE: Lecinski, 2011, p.16-17

1.3. DEVELOPING A DIGITAL MARKETING STRATEGY

The consumer journey became much more complex due to digitalization and technology evolution, because the individual is now influenced by multiple online touchpoints. The purchase decisions in the online marketplace, that is the combination of different kinds of online presence, are affected by a variety of information, including search results, business websites, conversations on social media, and, most of all, customer reviews.

A well-defined strategy provides a clear vision on how the firm should compete, that's why it is essential to a company's success. Before developing a digital marketing strategy, it is important for the firm to make some considerations with a critical eye about the company itself (i.e. are my products/services appropriate for online promotion? Are my business processes ready to include digital marketing?), the environment in which it operates (who are my competitors in the digital market? How does my targeted consumer use digital technology?) and its strategic priorities (by setting S.M.A.R.T.[1] objectives).

It is possible to track everything that happens online in digital marketing, and businesses can compare their progress against predefined goals and key performance indicators (KPIs) (Ryan D., 2016). KPIs must provide the information

---

[1] Acronym for Specific, Measurable, Achievable, Relevant, and Time-based.

and answers that the business needs (Marr, 2012). The mechanisms through which a strategy is put into action are known as tactics (Olson et al., 2020). Digital marketing tactics' success changes according to each situation.

Usually, the priorities related to digital marketing involve increasing sales leads, as well as developing customer engagement, improving brand awareness, generating more traffic on the website, and improving web user experience. The following are the seven most common digital marketing tactics:

- Content Marketing, which includes the set of techniques useful to create and share quality contents to create engagement and lead users through the purchasing process.
- Search Engine Optimization (SEO), which is the practice of taking actions to improve a website's ranking on a search engine page.
- Email Marketing, that is a type of marketing that focuses on providing timely and appropriate information related to new products, discounts and other services via email to prospects and customers.
- Search and social ads, which is the advertising that appears on social networks and web pages when specific keywords or phrases are typed into a search engine.
- Data-driven personalization that consists in making marketing decisions relying on data about individuals instead of past choices.

- Marketing technology usage, that is essential to automate and speed marketing operations, gather and analyze data, and give a variety of ways to reach and engage the targeted audience.
- Social media advertising, which refers to paid advertising campaigns on social media platforms in order to reach the target audience.

It would be simple to spend substantially in all these seven digital marketing tactics if marketing funds were infinite. Marketers need to determine how to effectively spend the company's promotional funds. According to a study carried out by Eric M. Olson, Kai M. Olson, Andrew J. Czaplewski, and Thomas Martin Key, "it appears that a marketing manager cannot go wrong by investing in content marketing and search engine optimization" (Olson et al., 2020, p. 292). Indeed, they seem to be popular among firms, implying that they should receive the largest share of a marketing budget.

Companies are present on numerous social media platforms for social media marketing, such as Facebook, Instagram, and TikTok in order to engage with their audience to build their brand, increase sales, and drive website traffic. The choice of platforms mainly depends on their target audience and marketing strategy (Dwivedi et al., 2020).

Furthermore, social media's impact on CRM manifests itself at three levels: acquisition, retention, and termination of relationships with customers (Lamrhari et al., 2021). Indeed, since social media have a strong influence on customers, many businesses use them to increase *customer acquisition* by running compelling brand campaigns. Through brand fan pages companies can even boost brand popularity, generating a significant impact on consumers' behaviors and decisions, and they can also include social media into their attempts to *retain* and develop long-term customer relationships. The more customers are retained (spending time on the website, mobile app or social media page), the greater the chance of purchase.

Finally, companies may use consumer data from social media to take preventative measures and enhance their ability to detect customers who are more likely to terminate a relationship with the company itself. Such information may also be used in churn prediction models. In the event of a business-initiated termination, the firm is less likely to try to retain a churning customer since it has already made the choice to terminate the relationship with them, for example because that specific customer is not profitable.

1.4. THE DIGITAL CONSUMER

In recent years, customers' perception of the Internet has changed significantly. Consumers are more informed than before due to digital technology, which enables them to make online research, so that they are more aware and conscious when making purchasing decisions. Indeed, consumer behavior has changed dramatically as a result of technology advancements and the widespread use of mobile devices, which has had a direct impact on how people connect and utilize social commerce to make decisions and purchase online (Dwivedi et al., 2020).

Marketers' aim is to provide relevant content that will aid in the consumer's decision-making process and encourage them to buy the company's products or services. Therefore, it is fundamental to understand and analyze the digital consumer behavior. Digital consumer behavior analysis is a type of digital marketing research that looks at how people utilize digital channels to choose items and services to buy online. In this case, the aim of marketers is to understand customer insight such as their needs, features, and behavior. As a result of this analysis, companies will build a customer segmentation aiming at creating targeted methods to reach customers in the most effective way.

Other necessary aspects that companies should be aware of are related to the demand analysis, which entails factors such as the current Internet trends and levels of use, as well as the many online services available, and how they connect to the products and services that the company intends to offer online. One of the advantages of this type of study is that organizations can find chances for the exact purpose of influencing consumers and selling online according to the target's real use of digital media. These factors also help develop more effective digital marketing campaigns and set more realistic objectives while formulating a marketing strategy. Marketers must consider a variety of technological issues while developing a digital marketing plan. For instance, they should take into account the customers' access to digital platforms and level of use, the kind of technology used by suppliers and intermediaries, and, last but not least, competitors' communication strategy.

The practice of evaluating the factors that influence customers' actual utilization of digital services to determine the most effective method or campaign for reaching them is known as *demand analysis*. Tools for online marketplace analysis such as SimilarWeb, Nielsen, Comscore, and Google Analytics, help estimate the number of individuals looking for information as well as the popularity of various sorts of websites based on the number of unique visitors[2].

Demand analysis can help save time and money by reducing uncertainty and increasing the effectiveness of digital marketing communications campaigns (Chaffey and Ellis-Chadwick, 2016). By combining these tools to primary research, companies can develop a solid understanding of their target consumers' behavior, needs, and wants, as well as their tendency to use digital channels. Once this

---

[2] Individual visitors to a site measured through cookies or IP addresses on an individual computer. (Chaffey and Ellis-Chadwick, 2016)

information is known, digital marketers start examining how to attain marketing goals and aims with digital communication.

Web retailers invest significant resources to increase the *conversion rate*, that is the percentage of website visitors that complete a purchase (Di Fatta et al., 2017). With conversion marketing[3] tactics, marketers use communication to convert as many potential customers as possible into actual customers and repeat customers.

Multiple forms of communication may be implemented depending on where the consumer is in the purchasing decision process. For instance, companies can focus on advertising to attract visitors at the beginning of the process and on sales promotion in the last stages. The most relevant ratios which determine the efficiency level of the digital marketing strategy are the awareness efficiency, attractability efficiency, engagement efficiency, and conversion efficiency.

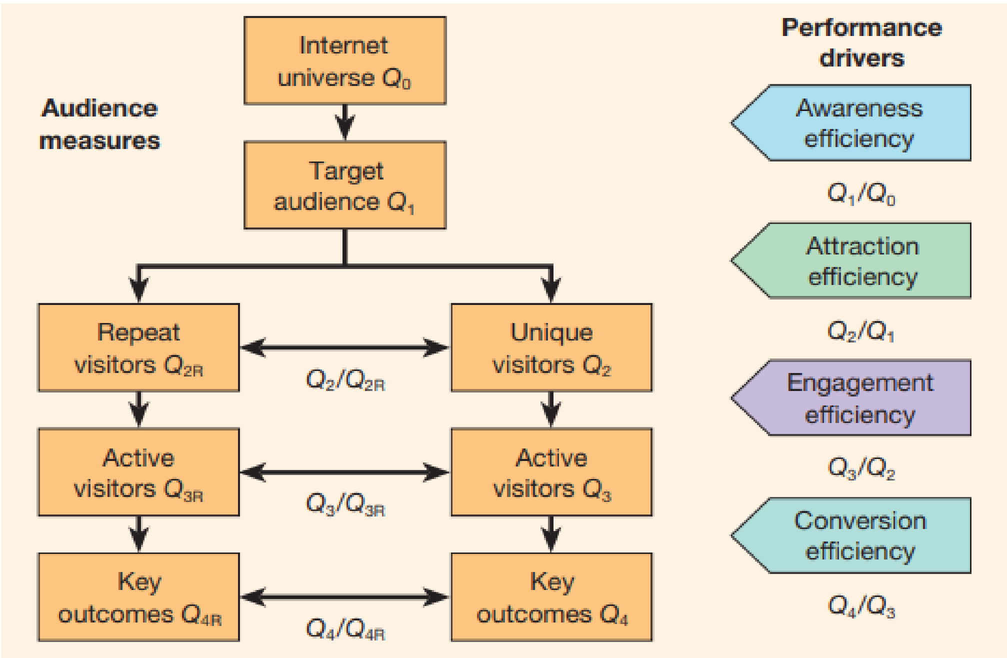
Another interesting ratio is the retention efficiency, which involves calculating the relationship between the number of repurchases/number of purchases (Chaffey and Ellis-Chadwick, 2016). This ratio helps understand whether the customers are satisfied by the products or services and the overall purchasing journey with the company. Clickstream for distinct segments is another essential aspect of online

---

[3] Using marketing communications to maximize conversion of potential customers to actual customers. (Chaffey and Ellis-Chadwick, 2016)

buyer behavior since it helps determine which pages a website visitor views before leaving.

Figure 1.4 - A model of the Internet marketing conversion process



A model of the Internet marketing conversion process. It shows key traffic or audience measures (Q0 to Q4), first-time visitors (Q2) and repeat visitors (Q2R) and key conversion efficiency ratios

SOURCE: Cheffy and Ellis-Chadwick, 2015, p. 70.

A company that has a high attractability efficiency but a poor conversion efficiency is clearly not performing well. Unfortunately, there is still a large percentage of businesses which focus more on creating appealing websites rather than on generating leads and sales to meet their marketing goals. Attracting potential customers by generating interest is essential to generate clicks and traffic, but the real challenge is to convert such traffic into sales. Companies may significantly enhance the returns on their advertising investments that generate traffic to a website by using conversion rate optimization (CRO) programs.

1.5. BIG DATA: THE REVOLUTION OF INFORMATION

Information is relevant not only for consumers, indeed, what customers find valuable is an evergreen question that firms ask themselves. During the last decades, it has been one of the most important questions related to marketing, and it is the reason why there has been a growing necessity to collect data and understand the individuals' needs and wants. The increased adoption of an omnichannel approach[4], which aims to improve shopping experience in both B2C and B2B settings, has also resulted in a rise in the quantity of data available to marketers (Marinelli et al., 2019).

---

[4] With an omnichannel approach, consumers can use all available channels and touchpoints interchangeably and fluidly switch between them (Verhoef et al., 2015; Mirsch et al., 2016)

Because of the technological revolution, which has made access to information simpler and faster, enterprises can use data in their decision-making process. The massive volume of high-speed data gathered by businesses to get valuable information and make crucial decisions is referred to as "big data". There are five features – also known as 5Vs - that define big data: volume, velocity, variety, veracity, and value. Enterprises manage extremely large databases (*volume*) whose data must be available on a real-time basis (*velocity*), can be obtained from many different sources (*variety*), and need to be clean and accurate (*veracity*) to be able to know customers and monetize all the information extracted by creating a competitive advantage (*value)* (Bumblauskas et al., 2017; Wang and Alexander, 2015; Zanoon et al., 2017). The fifth feature, value, is also the reason why data can have a quantifiable economic value (Marinelli et al., 2019).

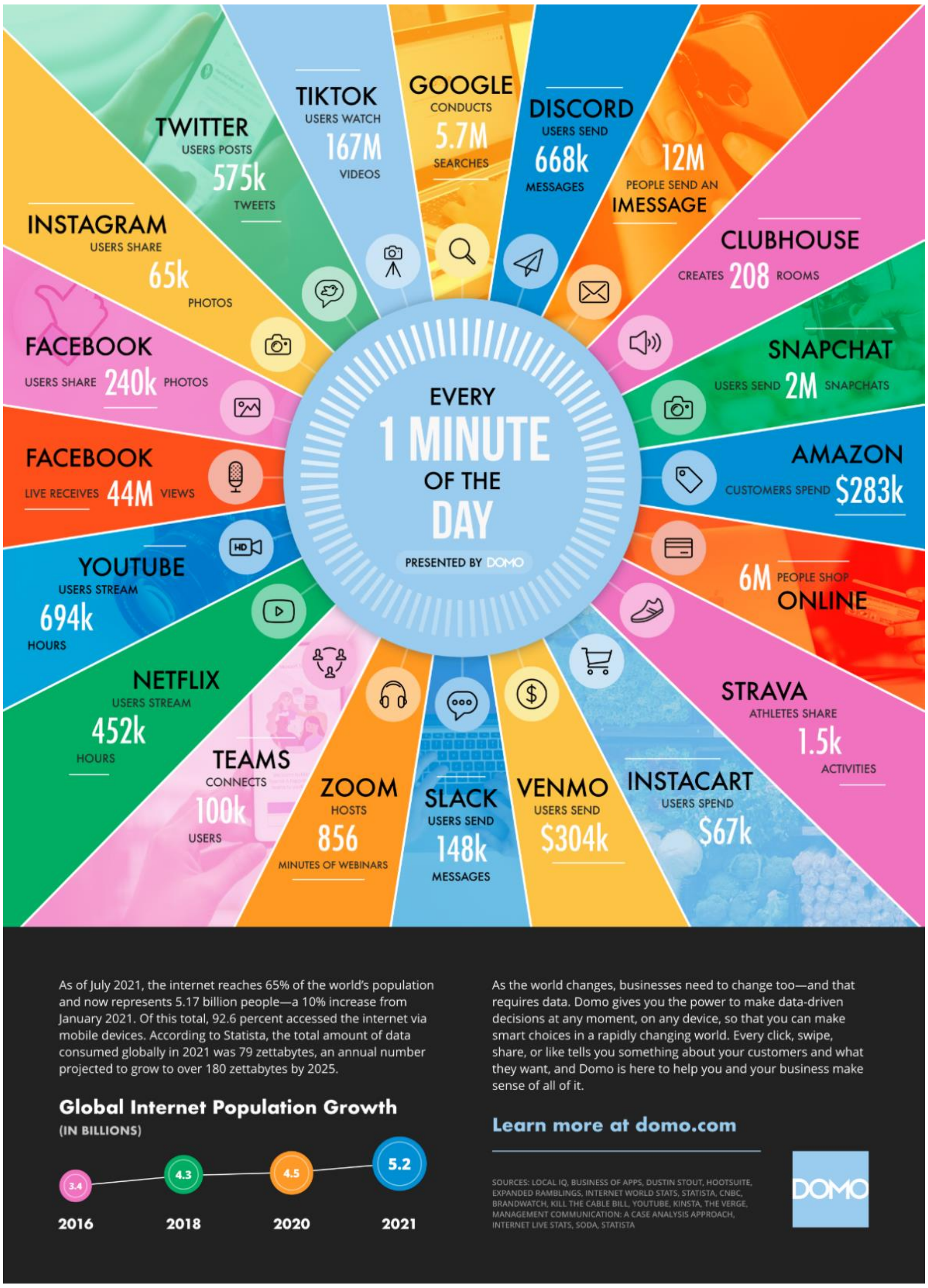
Big data is a cutting-edge technology that allows businesses to collect ever-increasing amounts of data about their consumers. Data originates from many sources, such as social interactions through user-generated contents (UCG), financial transactions in e-commerce websites, and non-financial transactions (Mariani et al., 2020). Big Data Management (BDM) and Big Data Analytics (BDA) are the two primary categories of big data procedures, though they cover different aspects of big data. Indeed, the former is a prerequisite of the latter, since BDA, which focuses on interpretation and knowledge discovery, requires data collection, pre-processing, storage, and sharing, all of which are carried out in BDM (Qi, 2020).

The advancement of technology, which resulted in the introduction of mobile devices such as smartphones and tablets, as well as improvements in mobile networks and WiFi, has resulted in an ever-increasing amount of data being created and consumed.

Every year, the cloud software company Domo, Inc., which specializes in business intelligence tools and data visualization, releases "*Data Never Sleeps*", an annual report that examines how users consume, generate, and display data every minute across high-traffic platforms and applications. According to the company, the Covid-19 pandemic has had a significant impact on the digital world. Digitalization has accelerated to the point that the number of people using the Internet climbed by 10 percent in the first quarter of 2021, with 65 percent of the world population accessing the Internet. Figure 1.5 shows what happens in 60 seconds on the Internet in 2021. Numbers are impressive. For instance, every minute, 6 million people purchase online and consumers spend $283,000 on Amazon.

Massive amounts of data are transforming the way people do business, socialize, do research, and govern society. Data is collected on everything, at all times and in every location (van der Aalst et al., 2017).

Figure 1.5 - Infographic on what happens on the Internet in a minute (2021)



SOURCE: Domo - https://www.domo.com/learn/infographic/data-never-sleeps-9

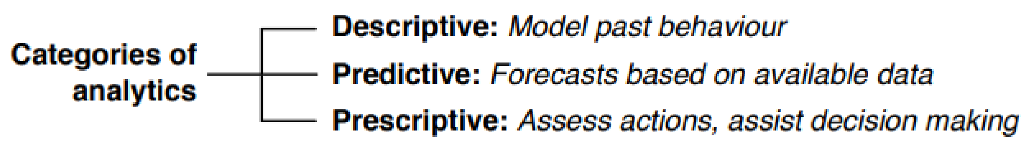
1.6. BIG DATA ANALYTICS: USES AND BENEFITS

The process of analyzing large datasets to disclose meaningful information such as market trends or customer preferences that may assist businesses in making more informed, better, and faster choices is known as Big Data Analytics (BDA).

BDA entails the use of specialized software to conduct statistical algorithms for prediction models and classification on high-performance platforms and aims to generate new insights that can meaningfully and, oftentimes in real time, complement traditional statistics, surveys, and archival data sources that remain largely static (Xiang et al., 2014). Through this process, businesses can get competitive advantage, since they are able to optimize inventories, find new revenue opportunities, improve customer satisfaction, define target customer segments, and so on.

The use of BDA involves several fields, from marketing to human resources. For instance, businesses may use data analytics in marketing to better understand the targeted segment and improve and customize their online advertising. It is also relevant for customer care, to predict customer churn, satisfaction, and behavior. Furthermore, BDA improves a business's supply chain management in strategic sourcing, demand forecasting, inventory and logistics. Last but not least, it helps in the Human Resources management by detecting the performance evaluation of employees. As shown in Figure 1.6, analytics solutions may be characterized as descriptive, predictive, or prescriptive.

Figure 1.6 – Descriptive, Predictive and Prescriptive Categories of Analytics



SOURCE: Assuncao et al., 2013, p. 5.

In particular:

- Descriptive analytics is concerned with understanding previous behavior and uses historical data to find trends and develop management reports;
- Predictive analytics, analyzes existing data as an attempt to forecast the future;
- Prescriptive analytics identifies actions and determines their influence on company's goals, requirements, and constraints to facilitate analysts in making decisions.

However, many businesses still lack expertise in data analysis and clear business objectives, making it impossible to obtain the necessary information for decision-making. Indeed, an excess of data often means a dead-end labyrinth. Therefore, only a few companies, the ones that make data their top priority, succeed in this challenge: in particular, Internet enterprises have made considerable use of big data technologies in their marketing management. Internet companies are those businesses that operate primarily via websites. By mining potential values of data and analyzing customer behavior and purchase records, it is possible to formulate efficient strategies to carry out the fundamental activities related to marketing. In order to reach a high position in the market, firms are increasingly relying on market data when developing product marketing mix strategies.

To take advantage of big data, marketing science will need to embrace disciplines such as machine learning, statistical learning, text-processing, audio-processing, and video-processing (Chintagunta et al., 2016). Data science is a field of study that extracts value from noisy, structured, and unstructured data and is based on scientific and statistical methods, processes, and algorithms. Machine learning, which is a branch of artificial intelligence (AI) and computer science, is helpful to make customized recommendations or detect unusual transactions. Moreover, machine learning uses text-processing to analyze, manipulate, and generate electronic text. Nowadays, text data is essential for enterprises to gain insights because the majority of customers' interactions with brands are online and text based. Indeed, as already mentioned, by analyzing text data, companies can understand how consumers search, purchase, and interact with them.

1.7. DATA-DRIVEN MARKETING

Data-driven marketing, that is the application of data and digital technologies in marketing, has supported the transformation and expansion of the scope of marketing. Data enables businesses to determine if marketing activities are performing well or not by defining valuable evaluation metrics, allowing managers to adopt the most effective marketing strategy.

Today's marketing decisions are strongly influenced by data analysis, which provides helpful insights into consumer behavior and preferences, as well as broader trends that affect marketing campaigns. Indeed, firms can track the effectiveness of each campaign in customer engagement, target users based on purchase history, and even predict consumer demand to run more efficient campaigns (Tripathi et al., 2021). Furthermore, big data analytics has acquired a lot of attention since it allows businesses to gain a competitive edge (Bumblauskas et al., 2017; Gupta et al., 2018).

With a prior analysis of valuable information gathered from complex and large data sets, companies are able to understand consumer demand and quickly formulate successful product marketing mix strategies. Data-driven marketing is essential to engage the right people, enable omnichannel experience, understand which are the best channels for promotion, understand customer behaviors, and figure-out what is working and what is not.

There are four main technologies that are expected to have a significant influence in marketing: the Internet of Things (IoT), Artificial Intelligence (AI), Machine Learning (LM), and Blockchain technology. More precisely, IoT may be necessary for data generation, AI and ML for data processing and analysis, and blockchain for security (Shah and Murthi, 2021).

Data analysis is perhaps the most important tool a company can use to implement customized e-commerce service. Data analysis tools may be adopted to fully improve the e-commerce marketing impact and build a tailored and accurate e-commerce marketing service system from three different perspectives: product marketing, channel marketing, and promotion marketing. In general, businesses promote their products and services through social media, which are then used by customers to receive information about such products and services. The buying behavior of consumers has changed a lot since the rise of the Internet and social media: they notice a product in a store and then purchase it online.

1.7.1. Digital Marketing Analytics to Analyze the Online Performance

Nowadays, the digital world generates such a large volume of unstructured data that it has been defined as a "data tsunami" (Sheth and Kellstadt, 2021). As a result, the importance of digital marketing analytics is constantly growing. The definition of web analytics - or digital analytics - provided by the Digital Analytics Association (DAA) claims that "*Web Analytics is the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage*".

The process of digital analytics includes both the tools that facilitate data collecting and the human analysts who actually use them in order to extract meaningful information. Many businesses are having difficulty with digital analytics because they are attempting to capture as much data as possible: they must be able to acquire and analyze only the data that can give valuable insights (Hemann and Burbary, 2018). Indeed, in order to ensure that opportunities do not fly under the radar, companies should assess their structure and investment in web analytics and digital marketing optimization (Chaffey and Patron, 2012).

Digital analytics involves the analysis of the performance of the following touchpoints (Bartoloni and Marinelli, 2019):

- Website, that can be a corporate website, a landing page, or an e-commerce. The analysis of a website is also known as *Web Analytics*;

- Online campaigns (display advertising and social media advertising);
- Email marketing campaigns
- Mobile applications
- Social Media.

As already mentioned in Chapter 1.3, businesses should define the objectives of the online activity they aim to achieve before choosing the Key Performance Indicators (KPIs). Indeed, KPIs are used to track whether or not the established goals have been achieved. A KPI for an e-commerce website, for example, may be the total number of orders placed by consumers divided by the total number of views.

Data collection for web browsing monitoring may be performed in a variety of techniques that have evolved over time. Log file analysis and page-tagging are the two most prevalent procedures (Bartoloni and Marinelli, 2019).

Log files represent the major source of data on the traffic system status, user behavior, system operations, and so on that is delivered via the web-node. Log files method collects information (IP address, country, search engine, average session duration, number of pages viewed, etc.) on the Web server and data is generated by devices, applications, operating systems, programmable or smart devices (Alguliyev et al., 2021; Bartoloni and Marinelli, 2019). On the other hand, *page tagging* collects data with invisible JavaScript code embedded in web pages and such data is then sent to a Web analytics external server, such as Google Analytics, Yahoo! Web Analytics, Adobe Analytics, and AWStats, to derive a variety of analytics. In general, page tagging-based tools are more widely used than transaction log file analysis-based tools as they are easier to use and less time-consuming (Nakatani and Chuang, 2011).

Cookies are another relevant element of these web analytics systems to track users and their behaviors. A cookie is a text string that is placed on clients' browsers as they connect to a given server (Cahn et al., 2016). The use of cookies has significant privacy implications for users. Collecting as much information on users as possible in order to deliver customized advertisements is clearly one of the main purposes of data brokerage.

1.7.2. Data Analysis for E-Commerce

With the increasing popularity of shopping websites, more and more people tend to purchase online goods and services and then make reviews. Though a single review might seem not useful, product review data contain highly valuable information for a firm.

Given the growing significance of e-commerce marketing, businesses must correctly understand consumers' consumption behavior and preferences based on marketing data analysis so that they can offer more accurate and tailored services and also reduce marketing costs. Indeed, big data and business analytics are used by businesses to develop effective and efficient strategies for the refinement of products and services to improve consumer satisfaction.

In the e-commerce world, 'big data' refers to the massive amounts of transactions, clickstream, audio, and video data that exist (Akter and Wamba, 2016). In e-commerce, big data can be gathered to build a *recommender system*. A recommender system is a set of algorithms that suggests items that could be of interest for a certain customer. It establishes a link between the user and the objects and exploits it to make recommendations. It helps in the search and selection of the appropriate product, promotes user engagement, and makes the contents more personalized. Recommendations can be based on:

- popularity, by suggesting items that have been viewed and purchased by a large number of people and also have a high rating. This method requires the use of descriptive statistics;
- classification, by using the features of each customer to decide whether they are interested in a product. It requires the use of classification methods such as logistic regression and tree-based methods to compute the probability of a customer liking some item;
- collaborative filtering, when the firm predicts the items that an individual might like based on the ratings given to that item by the other users with

similar tastes as the target user. Cluster analysis and matrix factorization are essential tools for its implementation.

In the e-commerce context, data analysis can show both customers' purchasing behavior and the industry's growth trend. This is fundamental for boosting the customized marketing effect. Big data and data analysis tools can assist businesses in gaining a better understanding of market dynamics so that they can establish a well-functioning customized marketing plan. The e-commerce's customized marketing strategy system based on data analysis mainly focuses on product, channel, and promotion strategy.

CHAPTER 2 – THE ROLE OF ONLINE CONSUMER REVIEW IN E-COMMERCE

2.1. DEFINITION AND USE OF OCRs

Nowadays, consumers search for product information online, where they can find product descriptions and experts' reviews. In addition to this, there is an increasing number of consumers who rely on other customers' evaluations of products, services, and brands to choose the product or service that best fits their needs (Salehan and Kim, 2016). Such evaluations are called online consumer reviews (OCRs) and are based on each individual's experience.

As shown by many studies (e.g., Bickart and Schindler, 2001), consumers prefer user-generated product information over the information provided by companies, since OCRs evaluate products and services from the user's perspective and seem to be more genuine: they are an electronic version of the word of mouth (WOM), also known as eWOM[5] (Electronic Word of Mouth). OCRs, indeed, have been demonstrated to positively affect both customer perceptions of the website's utility and social presence: they generate engagement, augment the time consumers spend

[5] An online extension of the traditional word-of-mouth: any positive or negative comment about a product, service or firm made by a potential, present, or past consumer that is published on the Internet and is available to a large number of individuals (Hennig-Thurau et al., 2004).

on the website, and develop a feeling of community among the habitual consumers (Mudambi and Schuff, 2010). As a matter of fact, by mutually sharing product knowledge and discussing purchasing experiences through OCRs, users may establish a sense of online community.

Many research (Riordan and Kreuz, 2010) highlight the strong influence of OCRs in consumers decision-making process, due to the fact that emotions may be efficiently transferred through computer-mediated communications (CMCs), influencing the way in which the receiver processes and interprets the message.

The number of disciplines involving the study of OCRs is huge and ranges from marketing to data science, statistics, psychology, and even information and computer science. In particular, data science focuses on the quantitative and qualitative properties and features of OCRs to interpret them through the usage of data mining, machine learning, and sentiment analysis techniques. Quantitative features concern ratings, volume or number of reviews, variance (that is the variation of the ratings), and also helpfulness (the number of users voting helpful the review), while qualitative features deal with the readability, sentiment, and informativeness of reviews. Reviewer and review quality, previous helpful votes, and the number of reviews already published are additional relevant factors to consider when evaluating the reliability and veracity of a review.

Each company gathers, processes, analyzes, reports and visualizes big data from OCRs differently. Businesses that use OCRs for BDA gain several insights and are able to improve their online activities. In particular, they can examine the customers' level of satisfaction with a product or service through rating and sentiment analysis, or even make navigation easier by allowing customers to filter reviews based on their level of usefulness. Perhaps, the greatest and most effective application of digital BDA is to assist new product development.

2.2. TRADITIONAL WORD-OF-MOUTH

Word-of-Mouth (WOM) was first characterized as face-to-face communication about items or companies between people who were not commercial organizations (Litvin et al., 2008).

Different consumers may post reviews for different reasons. Accordingly, Berger (2014) in his study, tries to investigate which are the behavioral drivers of word of mouth. Based on his research, WOM has five primary functions (see Figure 2.1): Impression Management, Emotion Regulation, Information Acquisition, Social Bonding, and Persuading Others. For each function, Berger identifies the "effect on sharing", namely the kind of information individuals should transfer with WOM. Moreover, a single occurrence of word-of-mouth might be motivated by numerous of these factors at the same time.

Figure 2.1 – The Five Functions of WOM and Their Effects

Function	Components	Effects On Sharing
Impression-Management	Self-Enhancement	+ Entertaining content + Useful information + Self-Concept relevant things + High status things + Unique and special things + Common ground + Accessible things + When aroused Shapes content valence
	Identity-Signaling	
	Filling Conversational Space	
Emotion Regulation	Generating Social Support	+ Emotional Content + Arousing Content Shapes content valence
	Venting	
	Facilitating Sense Making	
	Reducing Dissonance	
	Taking Vengeance	
	Encouraging Rehearsal	
Information Acquisition	Seeking Advice	+ Sharing when decisions are important or uncertain +Sharing when alternative info is unavailable or untrustworthy
	Resolving Problems	
Social Bonding	Reinforcing Shared Views	+ Common Ground Content + Emotional Content
	Reducing Loneliness and Social Exclusion	
Persuasion	Persuading Others	+ Polarized Content + Arousing Content

SOURCE: Berger, J. (2014), p. 589

For what it concerns *Impression-Management*, one reason customers spread WOM is to influence others' perceptions of them. Impression management is aided by communication in three ways:

- self-enhancement, because people like to be seen in a favorable light, thus they exhibit themselves in such a way by displaying their excellent purchasing decisions;
- identity-signaling, because people are willing to express their belonging to specific identities;
- filling conversational space by indulging in small chat, discussing nearly anything. For instance, when people run into someone they know, they may not have the purpose of saying the most intriguing thing imaginable, but they do not want to be silent.

Emotion regulation is the process of controlling the emotions individuals have, when they have them, and how they feel and express them. For instance, people may feel better if they vent their anger by sharing a negative experience they had, or they can engage in negative reviews to vent resentment and seek vengeance.

The most important function of WOM is perhaps *information acquisition*. WOM appears to help customers obtain knowledge by asking for advice and addressing difficulties. When consumers are uncertain on purchasing a product, they rely on

WOM for advice on what to do, how to cope with a particular problem and solve it, hear recommendations, or even simply get a different point of view.

Another function of WOM that might be obvious is to connect with other individuals (*social bonding*). Conversation is not exclusively limited to convey information: people have a basic need for social bonding and WOM helps fulfilling this human necessity.

Last but not least, WOM is used to *persuade others*. This appears to happen on a more interpersonal level as well as in a commercial situation. When people want to persuade others, they usually share more stimulating (e.g., anger-inducing or excitement-inducing) information contents.

2.3. FROM WOM TO eWOM

As technology has advanced and the Internet has become more widely used, opportunities for acquiring and disseminating product information have risen. Opinions are no longer communicated exclusively interpersonally (i.e., face-to-face), but rather via the use of information and communication technology (ICT) (Huete-Alcocer, 2017). Today, while socializing on social network sites (SNS), people exchange product evaluations and brand experiences in a variety of forms, including subjective statements, objective statements, and knowledge sharing

(Gvili and Levy, 2018). This event marked the introduction of a new term: eWOM (Electronic Word of Mouth). Though they present several differences, eWOM may be considered as a new wave of WOM communication. Similar to traditional word of mouth, eWOM is a kind of interpersonal communication transmitted by customers rather than marketers, giving it greater credibility and reliability. Nonetheless, the number of individuals involved in Internet communication, in terms of both contributors and audience, is much larger and there is the possibility of anonymity. In addition, time and geographical constraints are no longer an issue with eWOM, since they are available to other consumers for an indefinite period of time. Figure 2.2 illustrates the main differences between WOM and eWOM.

Figure 2.2 – Differences between WOM and eWOM

	WOM	eWOM
Credibility	The receiver of the information knows the communicator (positive influence on credibility)	Anonymity between the communicator and the receiver of the information (negative influence on credibility)
Privacy	The conversation is private, interpersonal (via dialogs), and conducted in real time	The shared information is not private and, because it is written down, can sometimes be viewed by anyone and at any time
Diffusion speed	Messages spread slowly. Users must be present when the information is being shared	Messages are conveyed more quickly between users and, via the Internet, can be conveyed at any time
Accessibility	Less accessible	Easily accessible

SOURCE: Huete-Alcocer, 2017, p. 3

Because of its easy accessibility, users consider eWOM appealing, and it has therefore become a popular source of consumer recommendation (Cheung et al., 2009). With eWOM, reviewers can share user-oriented product information as well as their personal recommendations through subjective and honest opinions about the product taken into account. The reviewers act as recommenders and provide aid to other users in making purchasing decisions. Overall, they facilitate other users in determining the quality of a product (Quambusch, 2015).

EWOM, however, not only helps increasing sales of a company. Another significant benefit of eWOM for a company is that it helps building and boosting costumer engagement [6](Sashi, 2012). Indeed, online customers can connect with a company by posting OCRs on Web sites, blog communities, and social media platforms. Engaged customers can be classified into passive, when their activities mainly involve the consumption of brand-related information, and active, when whey actively share brand-related information (Gvili and Levy, 2018). As a result, consumer eWOM engagement is defined as a two-dimensional construct that includes consumer intent to provide and receive eWOM. Getting customers to engage and participate in brand-related activities is one of the most difficult challenges marketers have to face when striving to establish strong and committed communities around the brand. In general, opinion-givers[7] are more likely to be technologically proficient, well-informed about product-related topics, and more willing to start conversations with others.

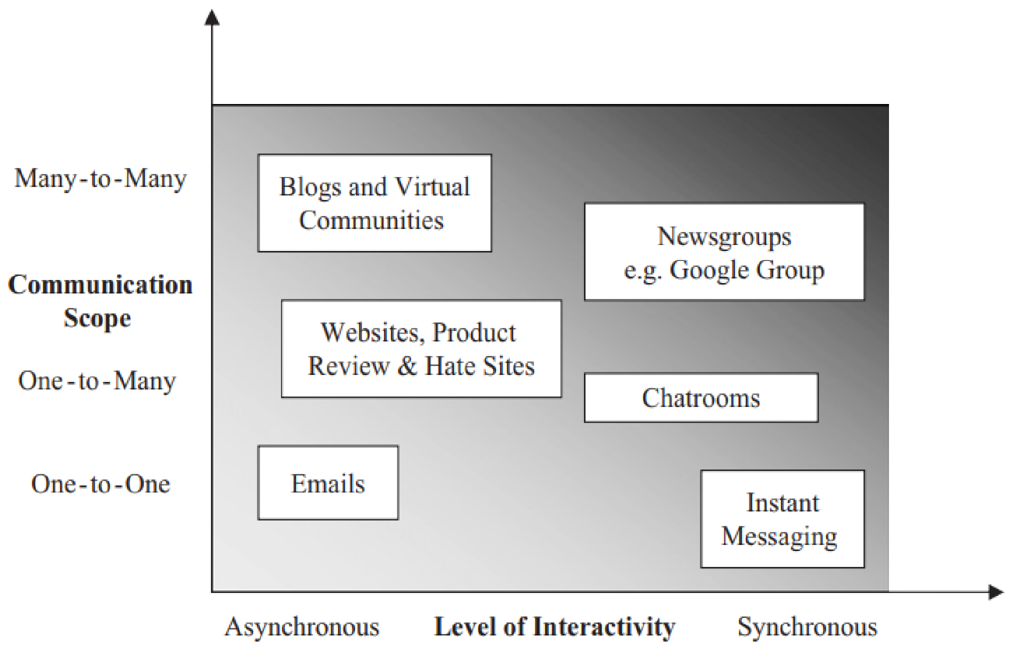
EWOM involves a variety of technological media that can be classified into different classes (Litvin et al., 2008) as shown in Figure 2.3:

[6] Customer behaviors that "go beyond transactions, and may be specifically defined as a customer's behavioral manifestations that have a brand or firm focus, beyond purchase, resulting from motivational drivers" (Van Doorn et al., 2010, p. 254, Gvili and Levy, 2018).
[7] Opinion-giving individuals are consumers who tend to actively generate and share brand-related information (Gvili and Levy, 2018).

- depending on the timing of communication, media can be synchronous (e.g. Instant Messaging) or asynchronous (e.g. email and blogs);
- according to the number of users involved, they can be One-to-One, when the communication links just one consumer with another (e.g. email), One-to-Many (e.g. web pages), or Many-to-Many (e.g. Internet chatrooms).

Unlike physical WOM, eWOM may develop virtual connections and communities.

Figure 2.3 - Electronic Word-of-Mouth (eWOM) Channels



SOURCE: Litvin et al., 2008, p. 462.

2.4. THE IMPACT OF ONLINE CONSUMERS REVIEWS ON SALES

Consumers frequently check online consumer reviews to make purchase decisions and perceive WOM and eWOM as a more valuable source of information than the one obtained through marketing channels. Moreover, as already stated, consumers frequently use the Internet to reduce the perceived risk of purchasing new items and services and are more willing to trust internet reviews (i.e., other people's opinions on a product) than suggestions from friends and relatives. As a matter of fact, when it comes to making decisions, customers are increasingly relying on eWOM to minimize risk: it tends to be more credible because the consumer who reviewed the product or service has experience with it (Huete-Alcocer, 2017).

Valence, volume, and dispersion of reviews are some of the most often utilized eWOM metrics (Chong et al., 2017). The *valence* of OCRs is defined as the evaluation score of a given product or service and has an indirect relationship with sales. Although it does not directly affect sales, a higher valence leads to a greater volume of reviews, which in turn increase consumer awareness and hence sales. Therefore, more eWOM discussions (volume) about a product or service will lead to higher customer awareness, which translates in sales increase. In addition, users of experience goods[8], since they are unable to experience product qualities, prefer

[8] Experience goods are those products or services that can be evaluated only after the purchase and experience (e.g. restaurant, hairdresser, beauty salon, theme park, travel, hotel).

to rely on extrinsic cues such as product popularity, which is shown in the volume of online reviews (Quambusch, 2015).

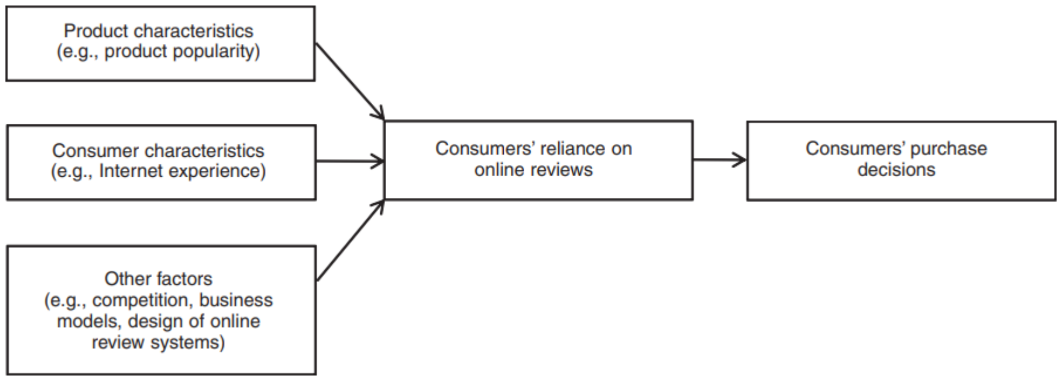
Trust is an essential requirement of eWOM. Trust is a source of credibility since it can push a user to make a purchase. The more a user trusts an online review, the higher the probability of purchasing a recommended item (and the lower the probability of purchasing a not recommended product). In general, a negative review has a major influence in the purchasing decision than positive feedback and it also spreads much faster than positive reviews (Chong et al., 2017). Furthermore, due to the high number of potential readers of eWOM communication and the protracted visibility of comments, a consumer's personal statement of a consumption problem can contribute to the shift of power from companies to consumers (Hennig-Thurau et al., 2004).

Previous studies have mostly focused on the impact of various review features on review usefulness, customer sentiments, and sales (Amblee and Bui, 2011; Purnawirawan et al., 2012) as well as their interactions with other parts of the product web page, such as pricing and product description (Maslowska et al., 2017). It is fundamental to understand which review elements draw customers' attention and, as a result, boost the influence of reviews in the decision-making process (Maslowska et al., 2020). Attention distribution, however, is a mental process that is generally difficult to analyze since it cannot be directly observed (Van der Lans et al. 2008).

Moreover, feedback from reviews has been found to have an impact on brand reputation (Davidson and Copulsky, 2006). Items belonging to a particular brand play a significant role in brand reputation, therefore, eWOM communication pertaining to the products in the portfolio is extremely likely to have a significant influence on the consumer's purchasing decision. Indeed, consumers are more willing to purchase a product of a brand they know for trust and loyalty reasons (Amblee and Bui, 2011).

Figure 2.4 portrays Zhu and Zhang's (2010) conceptual framework, according to which only when customers' trust on online reviews is sufficiently high the review is expected to affect product sales. The degree of trust is determined by product and consumer characteristics and other variables, such as competition, business models (e.g., business-to-consumer, consumer-to-consumer), or even the online review system's design (e.g. disposition of ratings, easiness of evaluating an item).

Figure 2.4 - Conceptual Framework of OCRs

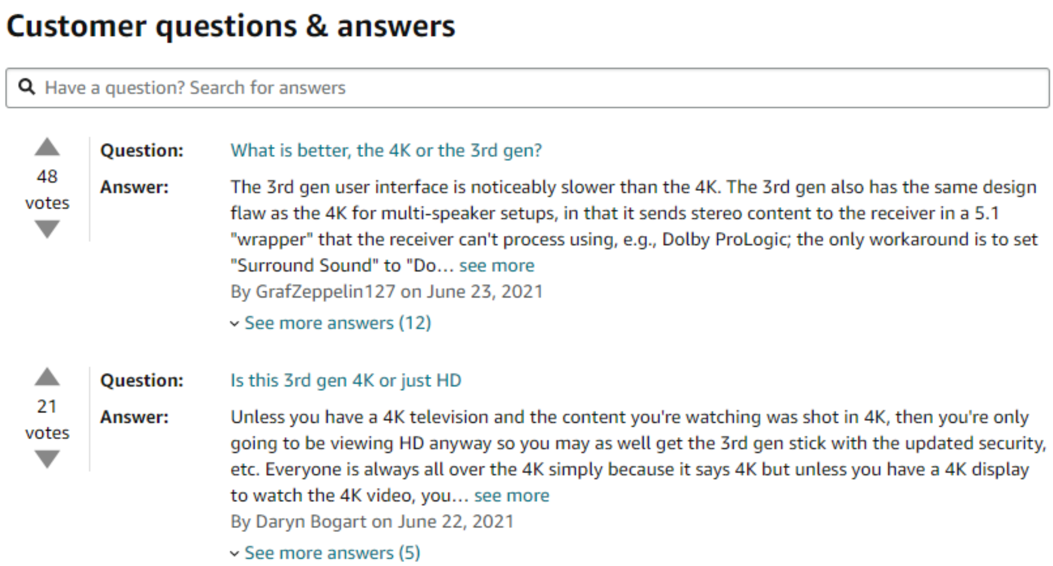


SOURCE: Zhu and Zhang, 2010, p. 135

In general, popular items' sales tend to be more influenced by OCRs, owing to the fact that popular products receive more reviews, and, as already mentioned, having a big number of reviews makes OCRs appear more trustworthy: the higher the number of reviews, the more accurate the overall rating. Moreover, since consumers are more certain that they can easily access online reviews of popular products, they are more inclined to look for them rather than for less-known products. Hence, the impact of this kind of reviews is projected to grow as a result of disproportionately more searches. On the other hand, consumers who assume that evaluations of less popular items are scarce and difficult to come by, may not look for them at all. Therefore, reviews of less well-known items would have minimal influence on their purchasing decisions (Zhu and Zhang, 2010).

Last but not least, in addition to providing consumers with the helpfulness of reviews (e.g. “Was this review helpful?”), several online marketplaces assist their customers by offering answers to their queries. Indeed, consumers are able to inquire extra information from online retailers through computer-mediated communication tools such as instant messengers (see Figure 2.5), and this has been demonstrated to be successful (Chong et al., 2017).

Figure 2.5 – Amazon’s Q&A tool



SOURCE: Amazon.com - https://www.amazon.com/fire-tv-stick-with-3rd-gen-alexa-voice-remote/dp/B08C1W5N87/ref=sr_1_1?keywords=fire%2Btv%2Bstick&qid=1645376817&sr=8-1&th=1

2.5. FAKE ONLINE REVIEWS

As already said, OCRs have a huge impact on customers' buying decisions, as well as product reputations, sales volumes, and merchants' profits (Wu et al., 2020). For this reason, online sellers have been paying close attention to them, sometimes they even attempted to alter customer attitudes by publishing fake reviews.

Although it might be considered as the act of publishing non-authentic, misleading OCRs in order to maximize profits, a universal definition of what constitutes fraudulent reviews does not exist so far (Salehan and Kim, 2016). Fake online review means both phony positive comments about the firm itself and negative reviews about competitors for financial gains (Wu et al., 2020). Imposters who manipulate the system by posting fake reviews are called *opinion spammers* and their activities are known as *opinion spamming* (Mukherjee et al., 2013). They are a critical e-commerce topic, since they heavily influence consumers, businesses and, consequently, the market efficiency.

Spam reviews are a serious problem on the Internet, they currently account for a significant portion of all existing online reviews (from 16% to even 33.3%) and are difficult to be effectively detected by humans, since they are written to be perceived as authentic (Cardoso et al., 2018).

But who are the promulgators of fake online reviews? Fake reviews are typically posted by small online merchants with weak brands, poor ratings, and low-quality products (Mayzlin et al., 2014, Zhuang et al., 2018). This might be due to the fact that firms with less popular products or brands invest more on advertising compared with well-known companies, since well-known brands may imply that their buyers will get information on their superior products from other sources.

However, even companies with strong brands, high ratings and quality, and competitive advantages sometimes publish fake online reviews. As a matter of fact, there have also been several events that highlighted the significance of fraudulent reviews in the e-commerce industry. TripAdvisor, for instance, was discovered to be involved in the creation of fake online reviews in 2012: about 50 million OCRs could not be considered real. Samsung was fined $340,000 in 2013 for creating fake reviews about one of its competitors, HTC. Last but not least, Amazon sued 1114 unidentified users in 2015 for writing fake online reviews (Wu et al., 2020).

Fake reviews have a strong impact on the growth of online product reviews, as well as the market and society, since they damage market efficiency and have a detrimental impact on social welfare. In addition, while spam on email, blogs, and social media is easy to spot, spam reviews are highly difficult to detect.

2.6. CREDIBILITY OF ONLINE CUSTOMER REVIEWS

Online reviews are a major source of eWOM (Electronic Word of Mouth) communication, and they have grown in importance as a marketing tool since many consumers look for online evaluations before making a purchase. The tendency to evaluate products and services by seeking information from others is defined as CSII (Customer Susceptibility to Interpersonal Influence) (Park et al., 2011). Indeed, consumers that are particularly vulnerable to interpersonal influence have been demonstrated to be more influenced by others and WOM (Schroeder, 1996).

The degree to which one believes a review is believable or truthful is known as "review credibility" (Cheung et al., 2009, Clare et al., 2018). When it comes to eWOM, a reader who believes the product review is legitimate is more likely to learn from it and use it. On the contrary, if the review is considered less trustworthy, its impact will be diminished, and the reader will be less inclined to follow the advice. Therefore, the adoption of eWOM reviews will be influenced by their perceived legitimacy.

Expertise, integrity and source attractiveness are all aspects that contribute to the idea of credibility. However, in view of the fact that everyone may publish a review with minimal editorial oversight or consent, the abundance of material widely accessible on the Internet poses the issue of legitimacy. Anonymity, as well as lack of universal standards for posting reviews might have a negative influence on reviews' credibility. OCRs are distributed through the Internet, and they are frequently posted by strangers. As a result, there is always a doubt regarding the system's integrity and dependability. It is the receiver's opinion that determines the review's credibility, and it can be both subjective and objective. The former opinion (subjective) is affected by attractiveness or expertise, whereas the latter is influenced by information quality.

The elaboration likelihood model (ELM) plays an important role when analyzing the credibility of a review. It began as a joint project between Richard Petty and John Cacioppo in the mid-70s, while they were both graduate students at Ohio State University (Petty and Briñol, 2011). They provided a useful way to understand how individuals are persuaded.

The ELM is a model that helps in clarifying how customers rate the reliability of reviews based on the numerous features of online reviews. In summary, it gives a basis to understand how persuasive communications are processed by individuals. Since the impact of a communication is highly influenced by credibility, it is essential to identify how people decide what to believe (Wathen and Burkell, 2002). According to ELM, persuasive messages may be processed via two primary routes: the central route, which requires extensive elaboration, and the peripheral route, which requires less elaboration. When consumers read a comment via the central route, they focus on the issues raised in the message and assess the arguments' merits. In these situations, they will engage in more cognitive processing and invest greater effort in order to evaluate such comments (Cheung et al., 2012).

Studies (Goldsmith and Horowitz, 2006) show that the majority of consumers rely on the opinions of others online for many reasons, such as to reduce the risk of making a poor purchase, ensure lower prices, obtain knowledge quickly and easily, by accident (unplanned), because they are driven by off-line inputs such as television, to obtain pre-purchase information and, last but not least, because they believe that other consumers' information is more important than advertising (Hennig-Thurau and Walsh, 2003).

However, while the majority of individuals read OCRs, there is a frequently overlooked small segment of customers who do not read them. A study carried out in Australia (Camilleri, 2021) focused on the demographic, psychographic, and behavioral factors that predicted whether or not someone would utilize OCRs. It showed that non-users of OCRs were more likely to be male, older, less educated, less technologically proficient, and neurotic, and saw OCRs as untrustworthy and ineffective.

Two main reasons emerged on why such non-users avoided online consumer reviews in the first place: lack of trust in OCRs and a preference for other sources, mainly personal experience. For various reasons, it is critical to understand who these non-users of OCRs are and what motivates them. Indeed, non-users of OCRs may be a vulnerable category in need of greater help from consumer advocates. Furthermore, non-users represent a customer category that must be targeted through other ways by firms that are more focused on online brand management and they also constitute an unexplored market for review platforms, and the reasons behind their non-use might shape these platforms' approach.

Figure 2.6 - Reasons Why People Do Not Trust OCRs

SOURCE: Camilleri, 2021, p. 6.

There is a strong probability that the lack of trust in OCRs comes from a perception that OCRs were biased or fake due to lack of author identity and inability to verify that such OCRs are written by real people.

In order to reduce such problems, companies should focus on improving or implementing particular features in their e-commerce platform. In particular, to guarantee the reviewer transparency, firms should include a profile page for each reviewer, showing their average review score, as well as the total number of reviews written, membership duration, and geographic location. By displaying such information and prioritizing reviews from verified consumers, OCR platforms might help relieve trust issues.

2.7. HOW TO EXPLOIT THE VALUE OF OCR

Today's shorter product life cycles and the large availability of substitute products put pressure on manufacturers and retailers to sell their products in a shorter period of time. In addition to this, more and more businesses are basing their commerce online, therefore, companies are investing a large amount of their resources to market and advertise their products on the Internet as a result of these commercial demands (Chong et al., 2017).

OCRs represent an essential marketing tool. A large number of reviews leads to an overall rating that quickly highlights whether a company is reliable and is worth the investment or if customers should purchase a competitor's products. More positive feedback equals more sales opportunities, and firms should understand how to encourage their consumers to share as much feedback as possible to improve their work.

An e-commerce website with positive reviews has a strong advantage over its competitors who ignore this significant factor. Companies should give their customers the chance to read reviews received from other consumers before deciding whether to purchase or not a product or service. As a result, displaying online consumers reviews offers consumers the assurance that the service provided is trustworthy.

All the above considerations imply that reviews are a critical factor both for consumers and companies. But how can a firm increase the volume of its online costumer reviews? In general, there are different post-purchase methods to encourage consumers to submit a review.

A common post-purchase marketing approach might be sending thank-you emails to buyers to express gratitude for the purchase they made and inviting them to share a review. Such emails also show customers that the company recognizes their worth as individuals and are a strategic tool to generate engagement. In change of a review, firms may even grant a discount or give a reward. Alternatively, companies may send newsletters to their active users highlighting the business's successes in terms of the overall user evaluation: this will express a sense of belonging and improve the relationship between the online store and the buyer.

Though it might be an underrated activity, taking the time to educate customers about the importance of submitting a review may be extremely beneficial to a business. A firm can also work on the display of its e-commerce website, for instance by including a popup alerts explaining the importance and how to write a review, or through calls-to-action on confirmation pages and follow-up emails, or even by making the process of leaving a review easier and faster.

In case of negative reviews, a good reaction is always a great move for a business. Indeed, negative reviews are unpleasant for any business, but failing to reply suggests that the firm does not care about the comments of its customers. Letting customers understand that they are heard and that the firm is eager to fix the problems can make the difference for a business, especially for two reasons: it demonstrates that the firm cares about its customers and their experience with them and conveys potential customers confidence in the company.

Furthermore, taking care of the brand image is essential since it represents the customers' perception of the company. A firm's reputation and brand value are enhanced by a positive brand image. Reviews and feedback may be expressly written on the company website but can also occur in forums, blogs, social media, and other channels (see Figure 2.3 - Electronic Word-of-Mouth (eWOM) Channels). Therefore, it is critical to monitor the perceptions and preferences of customers or users interested in a company's business, both to prevent negative reviews about its work from appearing on the web and to ensure that users are satisfied. As claimed by Davidson and Copulsky (2006), firms that ignore consumer feedback risk incurring low ratings and maybe a sharp drop in reputation and sales.

CHAPTER 3 - A PRACTICAL APPROACH TO DATA ANALYSIS WITH TEXT ANALYTICS AND SENTIMENT ANALYSIS

3.1. TEXT ANALYTICS AND SENTIMENT ANALYSIS TO ANALYZE ONLINE CUSTOMER REVIEWS

Text analytics, also known as text mining, is an AI technology that extracts interesting patterns or information from textual databases and transforms it into business intelligence. The data elaborated are extracted from text sources such as social media posts, online customer reviews, feedback comments and survey responses. It is possible to identify trends and popular topics and themes by extracting key sentences and words.

For a deeper inspection, sentiment analysis is implemented, giving insight into the emotion of the words and determining whether an expression is positive, negative, or neutral and to what degree, so that there is a general understanding of the opinions and emotions expressed in texts. More precisely, sentiment analysis aims to extract opinions from reviews and then classify them according to whether they are positive or negative (Elmurngi and Gherbi, 2017).

In order to make a more accurate and effective analysis, text analytics and sentiment analysis should be carried out together. For instance, text analytics shows the trending topics, while sentiment analysis applied to the same content highlights if the comments are positive or negative. When we read a text, we use our knowledge of the emotional meaning of words to determine if the text is positive or negative, or even defined by a feeling along the lines of surprise or disgust. Text mining methods should be used to pragmatically approach the emotional content of text. Indeed, for OCRs, text analytics and sentiment analysis help extract and analyze emotional expressions as well as cognitive and sensory-related processes, to understand whether a comment is positive or negative (Rong et al., 2021).

For interpreting the emotional intent of words, there is a range of approaches and dictionaries available. Text mining tools can be used to approach the text's emotional content: numerous packages contain many sentiment lexicons, that are collections of words with positive or negative sentiment orientation.

In general, lexicons are based on single words, also known as unigrams, and include a large number of words, for which a particular score is assigned to for positive or negative sentiment, and also possibly emotions such as joy, anger, surprise and sadness (Silge and Robinson, 2017). Moreover, since these tools show both sentiment and words, it is possible to analyze the number of words that contribute to each sentiment and then plot them into a histogram or *word cloud*, where the frequency of each word in the text is depicted by the size and color of the font, so that it is easier and quicker to determine prominent term and their relative importance.

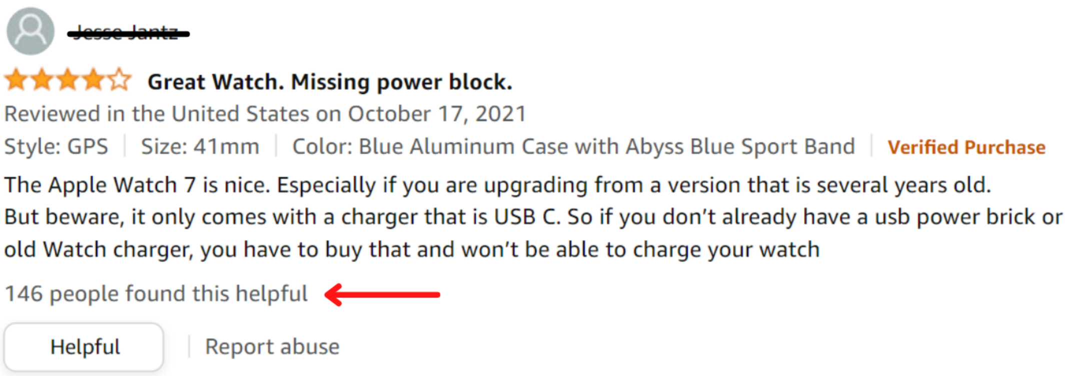
However, sometimes sentences may be structured in such a way that generates confusion in the analysis, because the single words may suggest a different sentiment than the entire sentence, and it is useful to look at different units of text. For instance, the statement “I am not having fun” is a sad sentence due to the negation, though, if analyzed in single words, it might be misleading. Therefore, there are sentiment analysis algorithms which try to comprehend the sentence’s sentiment as a whole by tokenizing text into sentences, rather than focusing on each single word.

There exist several measures to evaluate the relevance of OCRs. Sometimes, helpfulness, which may be considered as the value of the review, is used as the primary performance measure. Online retailers, such as Amazon.com, ask whether the review was helpful, so that they can show how many users thought the single review was useful and then the most useful ones are positioned at the top of the product’s reviews. There is also the numeric star rating that usually has a positive relation to sales, indeed, reviews which have many ratings do affect the sale of a product or service since they are considered to be more helpful. Another performance measure of OCR is the purchase intention, which is affected by the reviews' quantity and quality (Salehan and Kim, 2016).

Because the design of an online marketplace as well as how an online review is presented to the customer may be a significant predictor for online purchase intentions, a large percentage of online marketplaces have put a great amount of effort and resources into designing and improving the store interface (Van der Heijden and Verhagen, 2004).

A review interface usually consists of three basic elements: the review rating, the review content, and the review's helpfulness, which is displayed in a textual way such as: X people found this helpful (Korfiatis et al., 2012). An example of the review interface from Amazon US is provided in Figure 3.1.

Figure 3.1 - A review interface on Amazon.com



SOURCE: Amazon.com - https://www.amazon.com/dp/B09HF1DC1J/ref=fs_a_wt2_us1?th=1

This is the most helpful review of the Apple Watch Series 7 GPS on Amazon.com that the website puts at the top of the reviews. It shows the number of users who have found it helpful (146) and asks to contribute to voting whether it is helpful or not.

3.2. A PRACTICAL EXAMPLE OF TEXT ANALYTICS

3.2.1. Dataset information

This chapter aims at providing an example of how a firm which uses an e-commerce platform can apply text analytics to big OCRs data. The dataset used for this analysis is taken from Kaggle, an online platform of data science and machine learning, where users can find and publish datasets, work with data scientists and machine learning experts and participate in online challenges. The name of the dataset is "Women's E-Commerce Clothing Reviews" and includes 23,486 rows and 10 variables, which are:

- Clothing ID: refers to the reviewed item.
- Age: shows the age of the users who write the reviews.
- Title: refers to the title of each review.
- Review text: is the body of the review.
- Rating: refers to the score given to the purchased item by the user and goes from 1 (Negative Experience) to 5 (Positive Experience).
- Recommended IND: is a binary variable which indicates whether the user recommends (1) or not (0) the item.
- Positive Feedback Count: illustrates the number of users who voted the review as helpful.
- Division Name: refers to the high-level division of the product.
- Department Name: refers to the name of the product department.
- Class Name: refers to the name of the product class.

The data has been web-scraped from a real commercial firm and then anonymized by the publisher, indeed, references to the name of the business in the review text have been replaced with "retailer". The R package "quanteda" was used to carry out the analysis.

Once the "Recommended IND" variable has been transformed into a binary factor where "1" corresponds to positive feedback and "0" to negative, it has been possible to see that there are 3,575 negative and 16,087 positive reviews (see Figure 3.2).

Hence, two dataframes, respectively one with positive reviews and the other with negative ones, were created. Figure 3.3 provides an example of positive and negative review text.

Figure 3.2 – Table of Positive and Negative Reviews

```
> table(Reviews$Recommended)

    0     1
 3575 16087
```

SOURCE: own elaboration

Figure 3.3 – Positive and Negative Reviews Example

```
> Positive$`Review Text`[12]
[1] "Took a chance on this blouse and so glad i did. i wasn't crazy
 about how the blouse is photographed on the model. i paired it whit
 white pants and it worked perfectly. crisp and clean is how i would
 describe it. launders well. fits great. drape is perfect. wear tuck
ed in or out - can't go wrong."
> Negative$`Review Text`[15]
[1] "I was very excited to order this top in red xs. so cute, but it
 was huge, shapeless and support thin! it had to go back. i should'v
e looked at other reviews."
```

SOURCE: own elaboration

3.2.2. Text Mining and Overview of the "Quanteda" Package in R

Quanteda is a R package and is a basic tool for text analytics which allows researchers to gain useful insights from text. The package includes a set of functions for activities including *corpus*[9] management, tokenization, analysis, and visualization in natural language processing (Benoit et al., 2018). It also provides features for dictionary analysis, search for keywords in context, finding multi-word phrases using collocation score and so on.

Quanteda makes it straightforward to manage texts in the form of a corpus. Thanks to the package, users may easily split texts into words, paragraphs, sentences, or even delimiters and tags, organize them into bigger documents, and subset them using logical conditions or combinations of document-level variables (Benoit et al., 2015).

The documents in the corpus are usually modified with a pre-processing step in Text Mining, natural language processing (NLP), and information retrieval (IR).

- *Tokenization* is the process of separating a sequence of textual content into *tokens*, which can be words, phrases, symbols, or other significant items in

[9] A corpus is a set of text data that comprises document-level variables specific to each text. Two kinds of Corpus exist in R: the default implementation is the Volatile Corpus (VCorpus), whose term "volatile" refers to the fact that once the R object is destroyed, the entire corpus is lost, and the Permanent Corpus (PCorpus), which is a permanent object that can be stored outside of R (Feinerer, 2017).

order to investigate the meaning of the words in a sentence (Kannan et. al., 2014).

- *Stemming* is a commonly used technique in information retrieval activities for reducing a word to its stem or root form, improving precision and recall rate (Sharma and Cse, 2012). The purpose of stemming is to reduce a word's inflectional forms and derivationally related forms to a single base form (Jivani, 2011).
- *Stopwords*, also known as “noise words”, are words containing not relevant information for text analytics that appear frequently in documents. Examples of stopwords are determiners (e.g. “the”, “a”, “another”), coordinating conjunctions (such as “for”, “but”, “or”, “so”) and prepositions (e.g. “under”, “before”, “towards”). Because of their high frequency, stopwords can have an impact on the efficiency of the information retrieval process. Indeed, a list of stopwords is a crucial component of any machine learning activity requiring text processing (Kaur and Buttar, 2018).

3.2.3. Data Pre-Processing for Text Mining

Before starting text analytics, the dataset was cleaned by removing all the missing values and the variables “X1” and “Division Name” since they were not relevant for the analysis; as a consequence, the number of observations was reduced to

19,662. Approximately, reviews are 318 characters long, with a maximum of 508 characters and a minimum of 9.

The dataset also includes reviews containing special characters, such as the ampersand character (&), also known as “and sign”, exclamation marks and emoticons, as seen in Figure 3.4. The listed pre-processing steps in chapter 3.2.2. have been carried out with the aid of the “quanteda” package, as shown in Figure 3.5.

Figure 3.4 - Example of special characters in the reviews

```
> # ampersand character (&)
> Reviews$`Review Text`[126]
[1] "My usual size 6 fits perfectly... yes the metallic fibers on the inside are
 scratchy. a cami solves that problem. when ordering, i realized i cld not go st
rapless & wld have to wear a one of my wider strap cami's in navy or black to co
ver the bra strap area. it was obvious some sort of cover-up swtr or jacket w/b
 needed & wld cover the strap area anyway. so, i also ordered the 'faux-fur card
i' in the ivory to wear over this top. it's a shrug-like cardi w/ 3/4 slvs. come
s in the plum also if you"
> # multiple exclamation marks
> Reviews$`Review Text`[7461]
[1] "I was hesitant to buy this dress due to there not being any reviews, but de
cided to take a chance and order it. it is fabulous!!! great fit, off the should
er design stays in place, and the dress hits the smallest part of my waist perfe
ctly. the lace detail on the hem and the tassels on the ties are details that se
ts this dress apart. i wish that it was made in another color, perhaps a solid.
 you all hit this one out of the park, retailer!!!"
> # emoticons
> Reviews$`Review Text`[1857]
[1] "Ordered online and love it,exchanging however from s to m 'cos the fit is s
o perfect and i want to allow for extra room at the festive season:)"
```

SOURCE: own elaboration

Figure 3.5 - Data pre-processing for text mining

```
# Tokenize reviews
data.tokens <- quanteda::tokens(Reviews$`Review Text`, what= "word",
remove_numbers = TRUE, remove_punct = TRUE, remove_separators = TRUE,
remove_symbols = TRUE, split_hyphens = TRUE, remove_url = TRUE)

# Remove stopwords
data.tokens <- tokens_select(data.tokens, stopwords(),
                             selection = "remove")

# Stemming on the tokens
data.tokens = tokens_wordstem(data.tokens, language = "english")
```

SOURCE: own elaboration

After the pre-processing procedures, a document-frequency matrix (dfm) with a *bag-of-words model* may be constructed. A bag of words is a text representation used in Natural Language Processing (NLP) that describes how frequently a word occurs in a document. It simply focuses on whether or not a specific word appears in the document, not where it occurs (Great Learning, 2020).

Document-term frequency matrices have significant drawbacks. First, the longer the document, the higher the number of words. Second, not all the terms that recur often throughout the text are significant. In this specific case, *term frequency-inverse document frequency* (tf-idf) has been chosen as the main algorithm for document classification. Term frequency refers to the number of times a word appears in a document, while idf is the computation of the inverse probability of a term appearing in any document using the log of the inverse probability. Tf-idf is among the most relevant word weighting algorithms in text mining, and its accuracy in categorizing articles is promising, due to the fact that it examines the weight of each word by using two approaches: the occurrence of a word and the number of files in which a term may be located (Hakim et al., 2014). Figure 3.6 shows the R script for the creation of the document-frequency matrix and the weighting of such matrix by tf-idf algorithm.

Figure 3.6 - Creating and Weighting the Matrix by tf-idf

```
# Create a document-frequency matrix (bag-of-words model)
data.tokens.dfm = dfm(data.tokens, tolower = FALSE)


# Weight the matrix by tf-idf (now function dfm_tfidf)
data.tokens.dfm = dfm_tfidf(data.tokens.dfm)
```

SOURCE: own elaboration

3.2.4. Word Cloud

Popular subjects or trends may be detected using the DFM by plotting the most common terms: an example of such a plot is the word cloud. A word cloud is a graphic representation of the frequency of words in a text and it is useful for quickly determining prominent terms of a document and their relative importance. The properties of word clouds that have the greatest impact on users' attention are font size, weight, and color. Furthermore, words in the center of the cloud receive on average more attention than ones at the edges (Heimerl et al., 2014).

Word clouds are used in a variety of text analysis methods, such as patent analysis, opinion mining, and investigative analysis to graphically summarize text content.

3.2.5. Sentiment Analysis

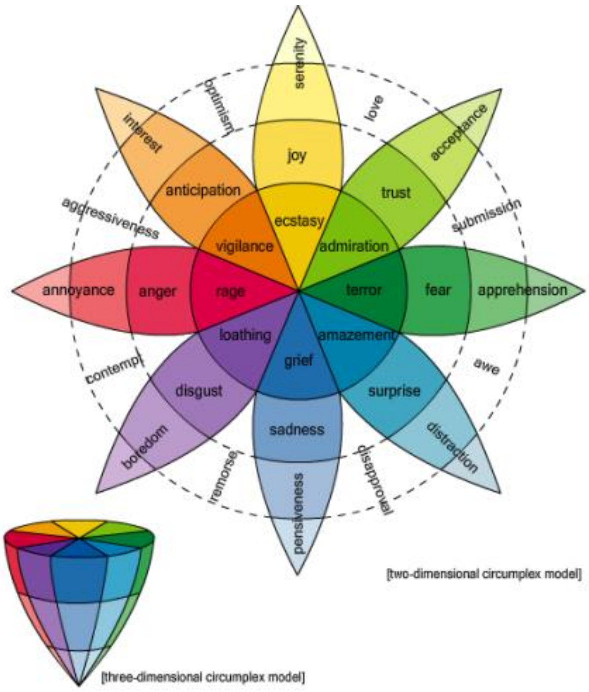
To understand the dominant emotions of words and whether the terms in the document are positive or negative, it is fundamental to use sentiment analysis, or opinion mining. The R package “syuzhet” has been used to carry out the analysis, which uses three primary lexicons: "AFINN," "Bing," and "NRC" lexicons.

A major focus was put on the NRC lexicon since, unlike Afinn and Bing, it presents is a list of 5,636 English words and their associations with eight basic emotions or sentiment columns (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) in addition to positive and negative sentiments. Therefore, it has been considered to be more accurate and provides a finer level of resolution.

Both the Afinn and Bing lexicons carry out a binary classification. Indeed, Bing only shows whether the sentiment is positive or negative and contains 2,005 positive words and 4,781 negative, for a total of 6,786 terms (Sonkin, 2021). Afinn is more precise, since it also values terms from -5 to 5, indicating the intensity of the sentiment from negative to positive (Hoyeol, 2018).

The ability to recognize emotions is beneficial for a variety of activities, including managing customer relationships by taking the necessary measures based on the emotional state of customers and tracking sentiment towards company’s products and services. Negative emotions, such as dissatisfaction and anger, can lead to complaints, litigations against the firm, unfavorable WOM, and other outcomes that are harmful to the organization's aims (Mohammad and Turney, 2013). Plutchik (1994), proposed a theory based on eight primary emotions, the same emotions adopted by the NRC lexicon, and organized them using a wheel whose radius symbolizes the strength of the emotion (see Figure 3.7). The closer is the sentiment to the center, the more intense it is. The eight basic emotions constitute four opposing pairs, respectively: joy–sadness, anger–fear, trust–disgust, and anticipation–surprise.

Figure 3.7 – Plutchick's Wheel of Emotions



SOURCE: Mohammad and Turney, 2013, p. 6

Figure 3.8 illustrates the nrc sentiment script in RStudio with the syuzhet package.

Figure 3.8 - RScript for nrc sentiment

```
sent = get_nrc_sentiment(text)
View(sent)
count.sent = colSums(sent)
```

SOURCE: own elaboration

3.3. MAIN RESULTS

3.3.1. Text Analysis Results

Before delving deeper into the text, a word cloud was plotted to have an overview of the most common words entailed in the reviews and to better comprehend the main topic of the feedback. Figure 3.9 shows a cloud of 200 words extracted from the dfm matrix of the example taken into account for this work. Some words of the word cloud seem to be cut. This is due to the stemming process that reduces words to their stem or root form to improve precision and recall rate. In this case, the most common words contained in the text seem to be related to the items, therefore, it implies that most reviews are about the item's quality rather than service. Moreover, positive adjectives tend to be more frequent, such as "love", "perfect", "great" and "nice": this is probably due to the higher number of positive reviews.

To understand whether both types of reviews are mostly focused on the items and their quality, two additional word clouds have been created for positive and negative feedbacks (Figure 3.10). As expected, also Negative reviews are about items.

Figure 3.9 - Word Cloud of 200 Words for Positive and Negative Reviews

SOURCE: own elaboration

Figure 3.10 – Positive Reviews Word Cloud VS. Negative Reviews Word Cloud

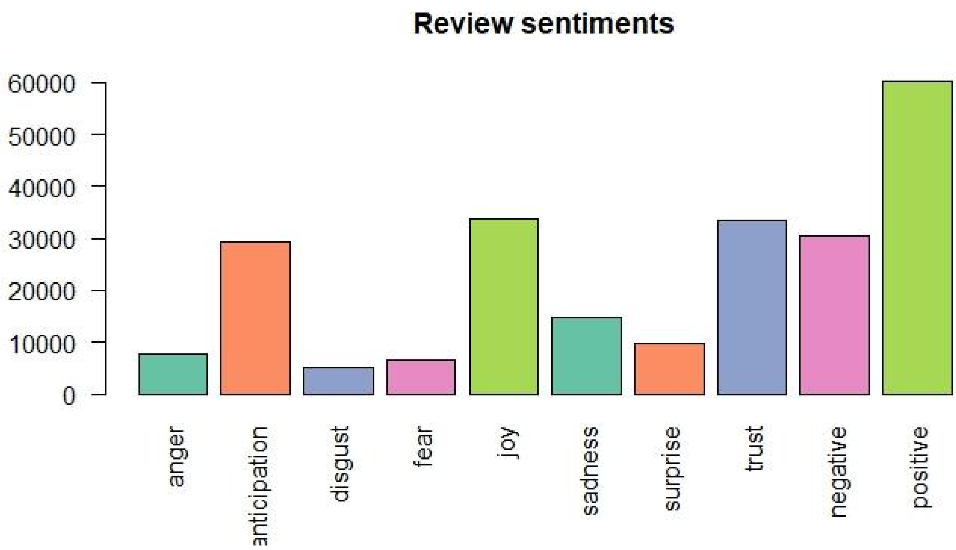
SOURCE: own elaboration

3.3.2. Sentiment Analysis Results

In order to show the different emotions of the text, barplot was created (see Figure 3.11). The vertical ax shows the number of times a single word associated with that sentiment occurs. In this case, there is a strong majority of words linked to positive sentiments (around 60,000 terms), such as trust, joy, and anticipation.

However, the strong presence of positive words does not imply that such words are related to positive reviews. Indeed, a deeper explanation might be given by n-grams, which contextualize words.

Figure 3.11 - Review sentiments



SOURCE: own elaboration

3.3.3. Understanding the Relationship Between Words With N-grams

Up to this point, major focus was put on the relationship between words (considered as separate entities) and sentiments or documents. Text analysis is also based on the correlation between words, for instance by looking at which words tend to co-occur inside the same document (Silge and Robinson, 2017). Thanks to *n-grams*[10], sequences of words are tokenized, rather than single ones. N-grams are useful because they provide a model of relationship between words by showing how frequently a given word is followed by another one.

For the present analysis, n to 3 was set, meaning that three consecutive words were grouped and examined. Figure 3.12 shows the R script used to carry out the n-gram process for the negative reviews of the e-commerce platform.

Figure 3.12 – Rscript for N-gram Function

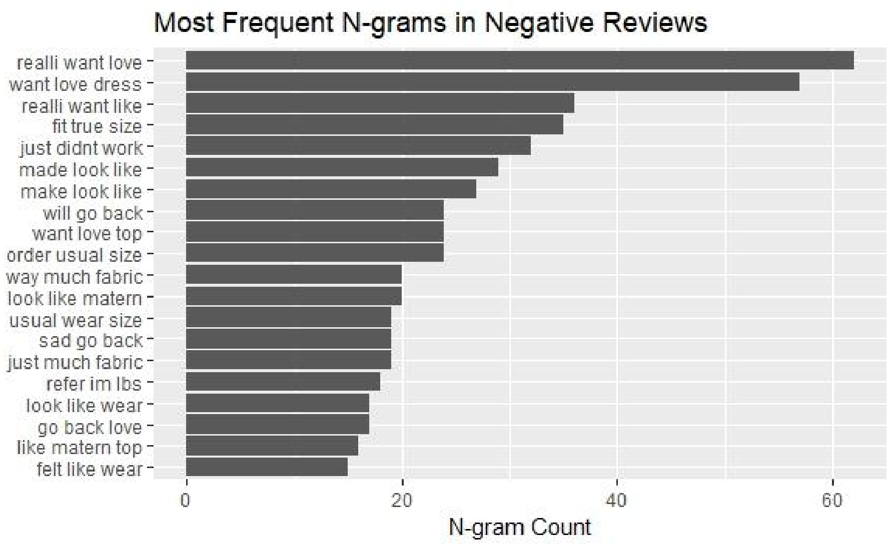
```
d = term_stats(Negative, ngrams=3)
# count is the number of appearances, and support is the number
# of documents containing the term
neg.bigram.freq = d[order(d$count, decreasing = TRUE),-c(3)]
```

SOURCE: own elaboration

[10] Function which tokenizes a sequence of adjacent words rather than individual ones (Silge and Robinson, 2017).

A list of the 20 most frequent 3-word-n-grams in negative reviews has been created (see Figure 3.13). The results might seem to be misleading, since they provide n-grams which seems to be positive such as "want love dress", "realli want love" and "fit true size". However, with the text_locate function it has been found out that such contexts referred to the disappointment of consumers who purchased the items. An example is given by Figure 3.14 and Figure 3.15, which analyze respectively the n-gram "realli want love" (62 observations) and "just didnt work" (32 total rows) and show the full text of two of the reviews.

Figure 3.13 – Most Frequent N-grams (n=3) in Negative Reviews



SOURCE: own elaboration

Figure 3.14 – Example of Negative Review With N-gram "Realli Want Love"

```
> text_locate(Negative, "realli want love", filter = NULL)
  text        before              instance          after
1 41                              realli want love  t excit receiv mai…
2 48   …ide look like tent  realli want love  top materi style s…
3 94   …rdigan cute pictur  realli want love  just didnt seem wo…
4 152                             realli want love  coat materi nice g…
5 211                             realli want love  dress someth fit f…
6 264  …eavi bulki flatter  realli want love  seem good qualiti …
7 382                             realli want love  blazer even tri se…
8 436  …size small perfect  realli want love  dress end return s…
9 471                             realli want love  dress just didnt f…

> Negative$`Review Text`[94]
[1] "This cardigan is very cute in the picture and i really wanted to
 love it but it just didn't seem worth the price. the material is kind
 of itchy to me and i wasn't a huge fan of the wide edges in the fron
t. the colour is very nice and feminine."
```

SOURCE: own elaboration

In the case of review number 94, some sentences or words may suggest that the review is positive (e.g. "the color is very nice and feminine"), nevertheless, the consumer was actually disappointed by the fabric of the item.

Figure 3.15 - Example of Negative Review With N-gram "Just Didn't Work"

```
> Negative$`Review Text`[628]
[1] "I ordered this blouse because it was such a good price on
 sale but should have paid closer attention to the reviews. the
 slits on the sleeves are much more noticeable than in the phot
o and the blouse just didn't work for me. something about it re
minded me of seinfeld's pirate shirt. i appreciate retailer's r
eturn policy. i took it back to the store and got an immediate
 refund on my cc."
```

SOURCE: own elaboration

Considering Figure 3.15 with review number 628 the customer is not satisfied with the item’s style, which appears to differ somewhat from the website's photographs. Another intriguing component of this specific review is the sentence “should have paid closer attention to the reviews”: this highlights how important reviews are for consumers and their influence on sales.

Overall, based on this analysis, the most common issues in negative reviews are related to the product's low quality and style of the cloths rather than to a scarce service. Indeed, the writer of review number 628 appreciated retailer’s return policy.

Because the Kaggle dataset does not include a pricing list for the goods, a sample of negative reviews mentioning price were taken into consideration in order to determine if the firm should remedy the quality problem by lowering prices or changing fabric suppliers (see Figure 3.16). It is clear that the price of such products is too high for their quality.

Figure 3.16 – Sample of Negative Reviews Referencing Price

```
> Negative$`Review Text`[18]
[1] "This is so thin and poor quality. especially for the price. it felt like a thin
 pajama top. the buttons are terrible little shell buttons. this could not have been
 returned faster."
> Negative$`Review Text`[32]
[1] "Bought this item on sale and was very disappointed in the quality for the cost.
 fabric feels cheap, like it will snag easily and will stretch out quickly. did not
 flatter the female form- felt like a burlap sack. gorgeous blue color but not worth
 the price tag. returned it."
> Negative$`Review Text`[231]
[1] "For the price of this dress, i was expecting something decent. the style is cut
e, but nothing more. however, the type of the polyester that the dress is made o
f.... oh my. it is a thick unpleasant material, that will cling to a body when it is
 hot. i never came across an item at retailer that was made this cheaply.\r\nthe tag
 on the back was hanging at half and the button on the front was ready to come off b
efore i even tried this dress on.\r\nno need to say - it's going back asap.\r\nyou s
hould not sel"
```

SOURCE: own elaboration

A further in-depth examination of the term "price" was conducted. Hence, a correlation analysis has been carried out to reveal meaningful relationships between different terms. It is interesting to check which words tend to co-occur with the word “price”. The findAssocs function has been used, and Figure 3.17 illustrates the results.

Figure 3.17 – findAssocs Function to Check “Price” Correlation

```
> findAssocs(Negative.dtm, "price", 0.2)
$price
outrag  worth
  0.38   0.30
```

SOURCE: own elaboration

The two words most correlated with the term “price” are “outrageous” and “worth”. This finding may imply that the price is too high, and the low quality of the items is not worth the price. However, this remains descriptive evidence, since the analysis is not inferential.

3.3. DISCUSSIONS AND MAIN IMPLICATIONS

The example demonstrated how e-commerce or website owners may examine their customers' reviews using text and sentiment analysis. As discussed in the previous sections, reviews are vital for businesses. Thus, it is essential to monitor and analyze data not only to manage the online reputation but also to get valuable insight into customers' perceptions and sentiments related to a company’s products or services. The example also proved that both positive and negative reviews’ analysis are essential to get information on the business’s performance.

The text analytics approach implemented for the analysis of the selected dataset has derived many insights from the OCRs. Each method, from the document-frequency matrix to the n-grams, contributed to information extraction. Overall, the reviews are positive, as evidenced by the “Recommended IND” binary variable, which revealed that the dataset had 3.575 negative reviews against 16.087 positive reviews at the beginning of the investigation.

The word cloud was beneficial for visually representing the frequency of words in OCRs so that prominent terms and relative relevance could be immediately determined. Two details emerged: the strong presence of positive words compared to negative ones and the connection to garments and their fabric. Such aspects provide useful information to the firm because, based on these results, the company can actually understand that its consumers appreciate its products and services. Hence, it is a positive outcome for the firm. The large number of positive reviews compared to negative ones and the connection to clothing and fabric were strengthened by the sentiment analysis and the n-grams. The former also revealed the occurrence of terms associated to sentiments of joy and trust.

However, it is also important to focus on the negative reviews since they can provide the company with useful insights. For example, they can shed light on which area needs to be improved in order to better meet and satisfy costumers' needs and lower the number of negative OCRs. For this deep inspection on negative reviews, the n-grams have played a significant role. A closer look at the negative reviews has shown consumers' disappointment and dissatisfaction with the fabric of the purchased clothing. Disappointment may be deduced from n-grams such as "want love dress", which shows that the customer had high expectations for the item but was disappointed when it turned out to be of inferior quality or damaged.

Furthermore, it emerged that the quality of items is not worth the price. It seems that prices are too high compared to the quality of the items. This result may be inferred from the correlation analysis as well as the random sample taken from the reviews filtered by the “price” word, where consumers complained about the high prices compared to the low quality of the fabric.

No review appears to be connected to the company's poor service, on the contrary, Figure 3.15 highlights a well-functioning return policy appreciated by the costumer.

Hence, to better satisfy the consumer needs and limit the negative reviews, the company should probably pay closer attention to price and quality. An attempt might be to reduce prices or improve the quality of the fabric.

CONCLUSIONS

Because today consumers interact with businesses via a variety of touchpoints across several channels and media, research continues long after a purchase is made, resulting in increasingly complex customer journeys. Digital marketing has led to new kinds of customer engagement and alternative models for information transmission that strongly differ from traditional marketing communications. Nowadays, consumers are more skeptical about the information provided by companies through advertising. Electronic Word-of-Mouth (eWOM), compared to traditional advertising, is growing in popularity among consumers and is thought to be more trustworthy than private signals.

The literature review confirmed a significant relationship between online consumer reviews and purchase intention. Volume, namely the number of reviews, contributes to raise customer awareness, which translates into sales increase. A negative review usually has a major influence in the purchasing decision than positive reviews and it also spreads faster.

However, to persuade a consumer to make a purchase, eWOM should be trustworthy. The stronger a user's confidence in an online review, the more likely they are to buy a suggested item, and vice versa. Product and customer characteristics, as well as other factors such as competition, business models, and even the display of the online review system, might influence the level of trust.

From a firm’s perspective, text analytics and sentiment analysis are crucial to monitor and understand data. Both positive and negative reviews provide companies with significant insights. It is clear that negative reviews are essential to identify which aspect needs to be adjusted to improve the whole performance of the company itself. Positive reviews, on the other hand, can disclose what customers like about a company, and they can help the company create a reputation and improve its placement.

The data analysis approach illustrated in Chapter 3 is only one example of how an e-commerce owner might examine its customers’ reviews. In fact, there exist numerous methods to analyze data depending on the information a company wants to learn. For instance, another analysis that could have been conducted with this dataset is a study of the reviews for each product category (i.e., jeans, dress, top, etc.) to see whether the negative reviews are mostly related to a specific product, or to identify the main complaints for each product category.

To conclude, for all the reasons stated above, a firm should always strive to increase the volume of its online costumer reviews. Encouraging customers to submit a review may be extremely useful to a company and there are several methods to do that. For instance, companies could send thank-you emails to express gratitude for the purchase and invite the buyers to share a review or grant a discount/reward in change of a review. Moreover, a firm's reputation and brand value are enhanced by a positive brand image. Thus, in case of negative reviews, a polite reply to such comments is usually an excellent business move.

One of the limitations of online consumer reviews' analysis might be sarcasm and disappointment, due to the fact that machine learning is presently unable to detect and extract such sentiment from reviews. To prevent this problem, a firm can add the option of rating the product (i.e. 1 to 5-star evaluation) to its e-commerce website, in addition to the review text, so that data analysts can separate positive and negative reviews and study them individually.

BIBLIOGRAPHIC REFERENCES

Alguliyev, R. M., Abdullayeva, F. J., & Ojagverdiyeva, S. S. (2021). Log-File Analysis to Identify Internet-addiction in Children. International Journal of Modern Education & Computer Science, 13(5).

Amblee, N., & Bui, T. (2011). Harnessing the influence of social proof in online shopping: The effect of electronic word of mouth on sales of digital microproducts. International journal of electronic commerce, 16(2), 91-114.

Askalidis, G., & Malthouse, E. C. (2016, September). The value of online customer reviews. In Proceedings of the 10th ACM Conference on Recommender Systems (pp. 155-158).

Assuncao, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., & Buyya, R. (2013). Big data computing and clouds: challenges, solutions, and future directions. arXiv preprint arXiv:1312.4722, 10.

Benoit, K., Nulty, P., Barberá, P., Watanabe, K., Lauderdale, B., Rcpp, L., & LazyData, T. R. U. E. (2015). Package ‘quanteda’.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. Journal of Open Source Software, 3(30), 774.

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. Journal of consumer psychology, 24(4), 586-607.

Cahn, A., Alfeld, S., Barford, P., & Muthukrishnan, S. (2016, April). An empirical study of web cookies. In Proceedings of the 25th international conference on world wide web (pp. 891-901).

Camilleri, A. R. (2021). Who doesn't read online consumer reviews, and why?. Personality and Individual Differences, 179, 110954.

Cardoso, E. F., Silva, R. M., & Almeida, T. A. (2018). Towards automatic filtering of fake reviews. Neurocomputing, 309, 106-116.

Chaffey, D., & Ellis-Chadwick, F. (2019). Digital marketing: strategy, implementation & practice. Pearson uk.

Chaffey, D., & Patron, M. (2012). From web analytics to digital marketing optimization: Increasing the commercial value of digital analytics. Journal of Direct, Data and Digital Marketing Practice, 14(1), 30-45.

Cheung, C. M. Y., Sia, C. L., & Kuan, K. K. (2012). Is this review believable? A study of factors affecting the credibility of online consumer reviews from an ELM perspective. Journal of the Association for Information Systems, 13(8), 2.

Cheung, M. Y., Luo, C., Sia, C. L., & Chen, H. (2009). Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. International journal of electronic commerce, 13(4), 9-38.

Chintagunta, P., Hanssens, D. M., & Hauser, J. R. (2016). Marketing science and big data. Marketing Science, 35(3), 341-342.

Chong, A. Y. L., Ch'ng, E., Liu, M. J., & Li, B. (2017). Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. International Journal of Production Research, 55(17), 5142-5156.

Clare, C. J., Wright, G., Sandiford, P., & Caceres, A. P. (2018). Why should I believe this? Deciphering the qualities of a credible online customer review. Journal of Marketing Communications, 24(8), 823-842.

Davidson, A., & Copulsky, J. (2006). Managing webmavens: relationships with sophisticated customers via the internet can transform marketing and speed innovation. Strategy & Leadership.

Di Fatta, D., Patton, D., & Viglia, G. (2018). The determinants of conversion rates in SME e-commerce websites. Journal of Retailing and Consumer Services, 41, 161-168.

Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., ... & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. International Journal of Information Management, 59, 102168.

Elmurngi, E., & Gherbi, A. (2017). Detecting fake reviews through sentiment analysis using machine learning techniques. IARIA/data analytics, 65-72.

Feinerer, I. (2017). Introduction to the tm Package Text Mining in R. Accessible en ligne: http://cran. r-project. org/web/packages/tm/vignettes/tm. pdf.

Felbermayr, A., & Nanopoulos, A. (2016). The role of emotions for the perceived usefulness in online customer reviews. Journal of Interactive Marketing, 36, 60-76.

García, M. D. M. R., García-Nieto, J., & Aldana-Montes, J. F. (2016). An ontology-based data integration approach for web analytics in e-commerce. Expert Systems with Applications, 63, 20-34.

Goldsmith, R. E., & Horowitz, D. (2006). Measuring motivations for online opinion seeking. Journal of interactive advertising, 6(2), 2-14.

Great Learning (2020), An Introduction to Bag of Words (BoW) | What is Bag of Words?, available at https://www.mygreatlearning.com/blog/bag-of-words/. Accessed January 28, 2022.

Gregori, G. L., & Pascucci, F., (2019). Il digital marketing come fattore competitivo. Verso un approccio integrato strumenti e strategia.

Gupta, S., Kar, A. K., Baabdullah, A., & Al-Khowaiter, W. A. (2018). Big data with cognitive computing: A review for the future. International Journal of Information Management, 42, 78-89.

Gvili, Y., & Levy, S. (2018). Consumer engagement with eWOM on social media: The role of social capital. Online Information Review.

Hakim, A. A., Erwin, A., Eng, K. I., Galinium, M., & Muliady, W. (2014, October). Automated document classification for news article in Bahasa Indonesia based on term frequency inverse document frequency (TF-IDF) approach. In 2014 6th international conference on information technology and electrical engineering (ICITEE) (pp. 1-4). IEEE.

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014, January). Word cloud explorer: Text analytics based on word clouds. In 2014 47th Hawaii international conference on system sciences (pp. 1833-1842). IEEE.

Hemann, C., & Burbary, K. (2018). Digital marketing analytics. Pearson.

Hennig-Thurau, T., & Walsh, G. (2003). Electronic word-of-mouth: Motives for and consequences of reading customer articulations on the Internet. International journal of electronic commerce, 8(2), 51-74.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronicword-of-mouth via consumer-opinion platforms: what motivates consumers toarticulate themselves on the internet?. Journal of Interactive Marketing, 18(1), 38-52.

Hoyeol (2018), *Limits of the Bing, AFINN, and NRC Lexicons with the Tidytext Package in R*, Digital Dialogue by Hoyeol Kim, https://hoyeolkim.wordpress.com/2018/02/25/the-limits-of-the-bing-afinn-and-nrc-lexicons-with-the-tidytext-package-in-r/. Accessed February 8, 2022.

Huete-Alcocer, N. (2017). A literature review of word of mouth and electronic word of mouth: Implications for consumer behavior. Frontiers in psychology, 8, 1256.

Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. Journal of King Saud University-Engineering Sciences, 30(4), 330-338.

Jim, L. (2011). ZMOT winning the zero moment of truth.

Jivani, A. G. (2011). A comparative study of stemming algorithms. Int. J. Comp. Tech. Appl, 2(6), 1930-1938.

Jockers, M. (2017). Package 'syuzhet'. URL: https://cran. r-project. org/web/packages/syuzhet.

Kannan, S., Gurusamy, V., Vijayarani, S., Ilamathi, J., Nithya, M., Kannan, S., & Gurusamy, V. (2014). Preprocessing techniques for text mining. International Journal of Computer Science & Communication Networks, 5(1), 7-16.

Kartajaya, H., Kotler, P., & Setiawan, I. (2016). Marketing 4.0: moving from traditional to digital. John Wiley & Sons.

Kaur, J., & Buttar, P. K. (2018). A systematic review on stopword removal algorithms. International Journal on Future Revolution in Computer Science & Communication Engineering, 4(4), 207-210.

Korfiatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. Electronic Commerce Research and Applications, 11(3), 205-217.

Lamrhari, S., El Ghazi, H., Oubrich, M., & El Faker, A. (2022). A social CRM analytic framework for improving customer retention, acquisition, and conversion. Technological Forecasting and Social Change, 174, 121275.

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

Mariani, M. M., & Wamba, S. F. (2020). Exploring how consumer goods companies innovate in the digital age: The role of big data analytics companies. Journal of Business Research, 121, 338-352.

Marr, B. (2012). Key Performance Indicators (KPI): The 75 measures every manager needs to know. Pearson UK.

Maslowska, E., Segijn, C. M., Vakeel, K. A., & Viswanathan, V. (2020). How consumers attend to online reviews: an eye-tracking and network analysis approach. International Journal of Advertising, 39(2), 282-306.

Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. American Economic Review, 104(8), 2421-55.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113.

Meyer, C., & Schwager, A. (2007). Understanding customer experience. Harvard business review, 85(2), 116.

Mirsch, T., Lehrer, C., & Jung, R. (2016). Transitioning to an omnichannel approach: A dynamic capability perspective. In 2016 International Conference on Information Systems, ICIS 2016. Association for Information Systems. AIS Electronic Library (AISeL).

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. Computational intelligence, 29(3), 436-465.

Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. MIS quarterly, 185-200.

Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013). Fake review detection: Classification and analysis of real and pseudo reviews. UIC-CS-03-2013. Technical Report.

Nakatani, K., & Chuang, T. T. (2011). A web analytics tool selection method: an analytical hierarchy process approach. Internet Research.

Olson, E. M., Olson, K. M., Czaplewski, A. J., & Key, T. M. (2021). Business strategy and the management of digital marketing. Business horizons, 64(2), 285-293.

Park, C., Wang, Y., Yao, Y., & Kang, Y. R. (2011). Factors influencing eWOM effects: Using experience, credibility, and susceptibility. International Journal of Social Science and Humanity, 1(1), 74.

Petty, R. E., & Briñol, P. (2011). The elaboration likelihood model. Handbook of theories of social psychology, 1, 224-245.

Plutchik, R. (1994). The psychology and biology of emotion. HarperCollins College Publishers.

Purnawirawan, N., De Pelsmacker, P., & Dens, N. (2012). Balance and sequence in online reviews: How perceived usefulness affects attitudes and intentions. Journal of interactive marketing, 26(4), 244-255.

Qi, C. C. (2020). Big data management in the mining industry. International Journal of Minerals, Metallurgy and Materials, 27(2), 131-139.

Quambusch, N. (2015). Online customer reviews and their perceived trustworthiness by consumers in relation to various influencing factors (Bachelor's thesis, University of Twente).

Riordan, M. A., & Kreuz, R. J. (2010). Emotion encoding and interpretation in computer-mediated communication: Reasons for use. Computers in human behavior, 26(6), 1667-1673.

Rong, L., Weibai, Z., & Debo, H. (2021, June). Sentiment Analysis of Ecommerce Product Review Data Based on Deep Learning. In 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (Vol. 4, pp. 65-68). IEEE.

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. Decision Support Systems, 81, 30-40.

Sari, P. K., Alamsyah, A., & Wibowo, S. (2018, March). Measuring e-Commerce service quality from online customer review using sentiment analysis. In Journal of Physics: Conference Series (Vol. 971, No. 1, p. 012053). IOP Publishing.

Sashi, C. M. (2012). Customer engagement, buyer-seller relationships, and social media. Management decision.

Schroeder, J. E. (1996). An analysis of the consumer susceptibility to interpersonal influence scale. Journal of social behavior and personality, 11(3), 585.

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. Information Sciences, 311, 18-38.

Shah, D., & Murthi, B. P. S. (2021). Marketing in a data-driven digital world: Implications for the role and scope of marketing. Journal of Business Research, 125, 772-779.

Sharma, D., & Cse, M. E. (2012). Stemming algorithms: a comparative study and their analysis. International Journal of Applied Information Systems, 4(3), 7-12.

Sheth, J., & Kellstadt, C. H. (2021). Next frontiers of research in data driven marketing: Will techniques keep up with data tsunami?. Journal of Business Research, 125, 780-784.

Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach. " O'Reilly Media, Inc.".

Sonkin, P. D., Sentiment Analysis of 49 years of Warren Buffett's Letters to Shareholders of Berkshire Hathaway, CRC Press, Bookdown, 2021

Tripathi, A., Bagga, T., Sharma, S., & Vishnoi, S. K. (2021, January). Big Data-Driven Marketing enabled Business Performance: A Conceptual Framework of Information, Strategy and Customer Lifetime Value. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 315-320). IEEE.

Van der Aalst, W. M., Bichler, M., & Heinzl, A. (2017). Responsible data science. Business & Information Systems Engineering, 59(5), 311-313.

Van der Lans, R., R. Pieters, and M. Wedel. 2008. Eye-movement analysis of search effectiveness. Journal of the American Statistical Association 103, no. 482: 452–61.

Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing: introduction to the special issue on multi-channel retailing. Journal of retailing, 91(2), 174-181.

Wang, L., & Alexander, C. A. (2015). Big data driven supply chain management and business administration. American Journal of Economics and Business Administration, 7(2), 60-67.

Wathen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the Web. Journal of the American society for information science and technology, 53(2), 134-144.

Wu, Y., Ngai, E. W., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. Decision Support Systems, 132, 113280.

Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction?. International Journal of Hospitality Management, 44, 120-130.

Zanoon, N., Al-Haj, A., & Khwaldeh, S. M. (2017). Cloud computing and big data is there a relation between the two: a study. International Journal of Applied Engineering Research, 12(17), 6970-6982.

Zhou, S., Qiao, Z., Du, Q., Wang, G. A., Fan, W., & Yan, X. (2018). Measuring customer agility from online reviews using big data text analytics. Journal of Management Information Systems, 35(2), 510-539.

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.

Zhuang, M., Cui, G., & Peng, L. (2018). Manufactured opinions: The effect of manipulating online product reviews. Journal of Business Research, 87, 24-35.

WEB DOCUMENTS

Dataset used for the practical approach. Retrieved from: https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews

https://wersm.com/what-happens-in-one-minute-online-2021-edition/

https://www.domo.com/learn/infographic/data-never-sleeps-9

https://nlp.stanford.edu/IR-book/html/htmledition/term-frequency-and-weighting-1.html

https://en.wikipedia.org/wiki/SEC_classification_of_goods_and_services

https://www.amazon.com/fire-tv-stick-with-3rd-gen-alexa-voice-remote/dp/B08C1W5N87/ref=sr_1_1?keywords=fire%2Btv%2Bstick&qid=1645376817&sr=8-1&th=1

https://www.amazon.com/dp/B09HF1DC1J/ref=fs_a_wt2_us1?th=1