# UNIVERSITÀ DEGLI STUDI DI PADOVA

## Dipartimento di Tecnica e Gestione dei Sistemi Industriali

Master's Degreee in Management Engineering

Master Thesis:

Sentiment Analysis - General Overview and Applications with focus on Social Media Platforms

Supervisor: Professor Scarso Enrico

Candidate: Lanaro Alex

School Year 2018 - 2019

**DICHIARAZIONE DI ORIGINALITA'**

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.

Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

**DECLARATION OF AUTHORSHIP**

The candidate declares that the present work is original and it is not been submitted yet, fully or partially, to the attainment of an academic degree in other Italian or foreign Universities.

Moreover, the candidates declare that all the materials consulted during the preparation of the paper are quoted in the text and in the section "References" and that eventual quotes are identifiable through the indication of the original publication.

Alex Lanaro

*To my biggest fan, my mum*

*To the man I want to be, my dad*

# Summary

The aim of this document is to analyze the concept of Sentiment Analysis, starting from its history to all the different approaches and methods you can encounter nowadays to its applications.

Concerning the applications, the main focus is related to social media, a relatively young sector who is now present in everyone’s life but it is still very far from the companies, especially the Small and Medium Enterprises.

It is first approached a brief overview of Social Media since the real essence and especially the most useful application of Sentiment Analysis is in this field.

Afterward, it follows the real body of the paper, with the Sentiment Analysis literature review, with different methods, approaches and obstacle that a company will face in case it decides to adopt sentiment analysis.

Last but not least, the paper ends with the analysis of the applications of a Sentiment Analysis tool in the business environment, first as general application following with real applications.

# Content Index

# 1. Introduction

We live in a world that is getting closer every day. Nowadays, Is possible to send a text message to a person on the other side of the planet in less than one second; to travel all over the world is affordable to everyone, and there are just 4 connections between us and Barack Obama [1].

Living in such a world can be both crazily scary or incredibly exciting, depending on how we approach the fast-changing that is taking place all around us.

If we should summarize the changing of today in one world it would be the internet. It is thanks to this amazing network that the world has known an unseen growth in term of speed and magnitude, an unstoppable transformation in every sector that affects the life of every human being and make possible things unimaginable before.

In the last decade, the word *connection* lost his previous meaning, acquiring an impact that goes behind the devices we have in our daily life, we are all connected at a level which goes beneath our skin.

Let’s imagine taking our mobile phone out from our pockets, unlock it, and give it to the person next to us. Scary, isn’t it? Why is it so scary? It is just a device we use to call and contact other people, right?

The reality, though, is way different. The connection level that our society reached at an incredible rate speed in the last years. With our phone, we can access all the people we know, our money in the bank account, see where we had been in the past trips, see our emails and everything we like, and we dislike.

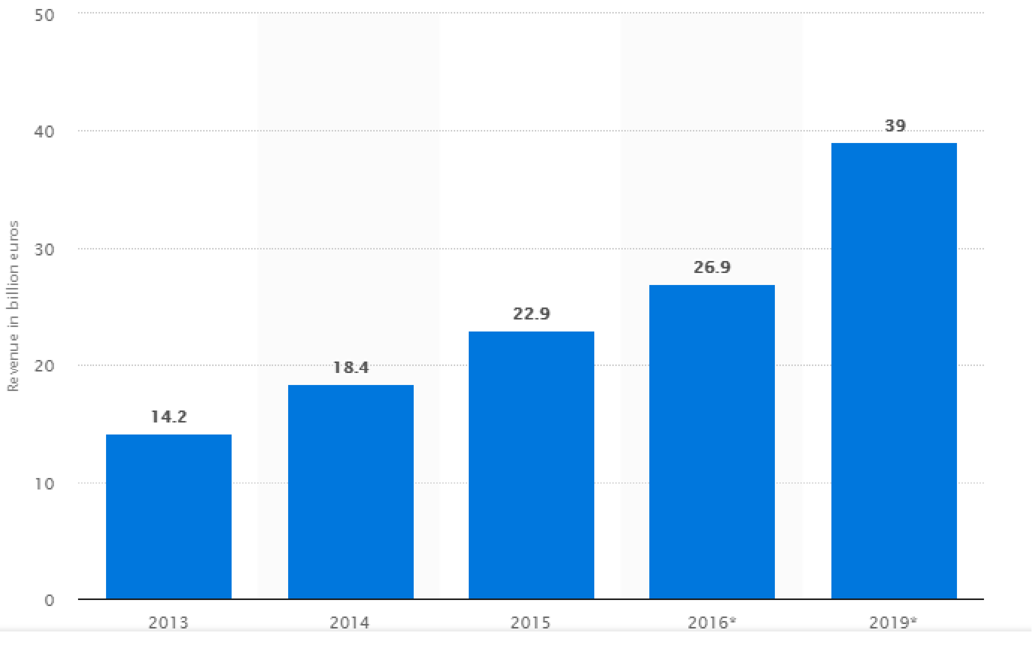
If we mention the internet and connection in the same sentence, it is no surprise if the first word that came into your mind is Social Media.

Social Media are the connection concept exaggerated at the maximum level. We can meet and talk with someone on the other edge of the planet, we can see what other people do, like, comment, where they are, with whom and why all of this without even getting up from our bed. The other side of the medal is that of course anyone can see our life as well, but this is the price to pay to be connected.

We can say so, without fear of been contradicted by anyone, that Social Media Platforms are shaping everything in the world.

In the last five years, the Social Media market has become huge. Online advertisement, Big Data, machine learning, cloud sharing, artificial intelligence, are just some of the industries born from the development and increase of the connection and of the role of social media in our society, creating a huge market out of nowhere, a market that 15 years ago was just some few dreamers' minds.

There is no sector, no business, no institution that cannot take this huge market into consideration. I say huge because the sector by itself has revenues in 2017 for more than 39 billion dollars [2] and is keep growing. It will be fouling to not try to get a slice of these big cake.




Tab A Revenues coming from Social Media Networks from 2013 to 2019

The problem is: how?

How to get a slice of this money? And is it actually about money?

No, is not anymore about money, not in the way we normally think about them anyway. I'm not talking about Apple Pay, I'm not even talking about Cryptocurrencies. I'm talking about the money of the future.

In the 70s, regarding the new advertisement transmitted in TV cable programs, popped out the quote "If you are not paying for the product, you are the product" [36]. It follows that, since we are not paying to use Facebook, Twitter, Instagram and so on, we are the product, and we are unquestionably paying. So, what are we paying for using all these platforms for free?

We are paying with data.

Data nowadays are so important that the tech giants (Apple, Facebook, Google, Amazon, Microsoft) want them at every cost.

Quoting the Economist "A new commodity spawns a lucrative, fast-growing industry, prompting antitrust regulators to step in to restrain those who control its flow. A century ago, the resource in question was oil. Now similar concerns are being raised by the giants that deal in data, the oil of the digital era". [30]

What has changed? Smartphones and the internet have made data abundant, ubiquitous and far more valuable.

Whether you are going for a run, watching TV or even just sitting in traffic, virtually every action creates a digital trace - the more raw material for the data machines available to each one of the above-mentioned tech companies.

Focusing the attention on the Social Media part, this abundance of data changes the nature of competition. Technology giants have always benefited from network effects: the more users Facebook signs up, the more attractive signing up becomes for others. With data, there are extra network effects.

By collecting more data, a firm has more scope to improve its products, which attracts more users, generating even more data, and so on, creating a loop which is hard to break.

The ways you can analyze those data are almost infinite, so I choose the one that I found almost magical because it gives the sensation to read on people's mind: The Sentiment Analysis,

Sentiment Analysis is a powerful structure that allows you to understand what people, millions of them, think, against who and why, giving you the power to predict the future, create the perfect product for your customers or, why not, predict who is going to win the next presidential elections.

That is why I choose to study Sentiment Analysis, to combine a market that is growing with a rate that has no comparison, with what can actually make the difference in the future, data or, to be more precise, data analysis.

"Data is a precious thing and will last longer than the systems themselves."

Tim Berners-Lee, father of the Worldwide Web

I hope that you can learn something by reading this paper and I hope to keep your interest up at least as much as I enjoyed the work. "

# 2. Social Media: a brief overview

## a. Definition

Before diving into the Sentiment Analysis and its application to the Social Media platforms, we have to understand them separately.

“We define social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.” [3]. The nature and name of these connections may vary from site to site.

What makes social network sites unique is not that they allow individuals to meet strangers, but rather that they enable users to articulate and make visible their social networks. This can result in connections between individuals that would not otherwise be made, but that is often not the goal, and these meetings are frequently between ‘‘latent ties’’ [4] who share some offline connection.

While SNSs have implemented a wide variety of technical features, their backbone consists of visible profiles that display an articulated list of Friends who are also users of the system. Profiles are unique pages where one can ‘‘type oneself into being’’ [5].

The visibility of a profile varies by site and according to user discretion. By default, profiles on Friendster and Tribe.net are searched by search engines, making them visible to anyone, regardless of whether or not the viewer has an account.

Alternatively, LinkedIn controls what a viewer may see based on whether she or he has a paid account. Sites like Myspace allow users to choose whether they want their profile to be public or ‘‘Friends only.’’ Facebook takes a different approach—by default, users who are part of the same ‘‘network’’ can view each other’s profiles, unless a profile owner has decided to deny permission to those in their network. Structural variations around visibility and access are one of the primary ways that SNSs differentiate themselves from each other.

After joining a social network site, users are prompted to identify others in the system with whom they have a relationship. Most SNSs require bi-directional confirmation for Friendship, but some do not.

The public display of connections is a crucial component of SNSs. The Friends list contains links to each Friend’s profile, enabling viewers to traverse the network graph by clicking through the Friends lists.

Most SNSs also provide a mechanism for users to leave messages on their Friends’ profiles. This feature typically involves leaving ‘‘comments,’’ although sites employ various labels for this feature. In addition, SNSs often have a private messaging feature similar to webmail.

### b. The Early Years

According to the definition above, the first recognizable social network site was launched in 1997.

SixDegrees.com allowed users to create profiles, list their Friends and, beginning in 1998, surf the Friends lists. Each of these features existed in some form before SixDegrees, of course. Profiles existed on most major dating sites and many community sites. AIM and ICQ buddy lists supported lists of Friends, although those Friends were not visible to others. Classmates.com allowed people to affiliate with their high school or college and surf the network for others who were also affiliated, but users could not create profiles or list Friends until years later. SixDegrees was the first to combine these features. [3]

SixDegrees promoted itself as a tool to help people connect with and send messages to others. While SixDegrees attracted millions of users, it failed to become a sustainable business and, in 2000, the service closed. Looking back, its founder believes that SixDegrees was simply ahead of its time.

After a number of communities created from 1997 to 2001, the next wave of SNSs began when Ryze.com was launched in 2001 to help people leverage their business networks. Ryze's founder reports that he first introduced the site to his friends.

In particular, the people behind Ryze, Tribe.net, LinkedIn, and Friendster was tightly entwined personally and professionally. They believed that they could support each other without competing [6].

In the end, Ryze never acquired mass popularity, Tribe.net grew to attract a passionate niche user base, LinkedIn became a powerful business service, and Friendster became the most significant, if only as ''one of the biggest disappointments in Internet history'' [7].

### c. SNSs take off

From 2003 onward, many new SNSs were launched, prompting social software analyst Clay Shirky [8] to coin the term YASNS: ''Yet Another Social Networking Service.'' Most took the form of profile-centric sites, trying to replicate the early success of Friendster or target specific demographics.

Furthermore, as the social media and user-generated content phenomena grew, websites focused on media sharing began implementing SNS features and becoming SNSs themselves. Examples include Flickr (photo sharing), Last.FM (music listening habits), and YouTube (video sharing).

With the huge number of venture-backed startups launching in Silicon Valley, few people paid attention to SNSs that gained popularity elsewhere, even those built by major corporations, for instance with the huge flop of Google's Orkut, the SNS launched by the Mountain Dew Company which never reached a sustainable user base in the U.S.

Few analysts or journalists noticed when MySpace was launched in Santa Monica, California, hundreds of miles from Silicon Valley. MySpace began in 2003 to compete with sites like Friendster, Xanga, and AsianAvenue, according to co-founder Tom Anderson.

MySpace growth rapidly thanks to the users of Friendster, who posted on the social media messages encouraging people to join the competitor.

While MySpace was not launched with bands in mind, they were welcomed. Indie-rock bands from the Los Angeles region began creating profiles, and local promoters used MySpace to advertise VIP passes for popular clubs.

Bands were not the sole source of MySpace growth, but the symbiotic relationship between bands and fans helped MySpace expand beyond former Friendster users. The bands-and-fans dynamic was mutually beneficial: Bands wanted to be able to contact fans, while fans desired attention from their favorite bands and used Friend connections to signal identity and affiliation.

Furthermore, MySpace differentiated itself by regularly adding features based on user demand [9] and by allowing users to personalize their pages. This ''feature'' emerged because MySpace did not restrict users from adding HTML into the forms that framed their profiles; a copy/paste code culture emerged on the web to support users in generating unique MySpace backgrounds and layouts [10].

Teenagers began joining MySpace in masse during 2004 encouraging friends and, after these years of exponential and uncontrolled growth, News Corporation purchased MySpace for $580 million in 2005 [11] attracting massive media attention.

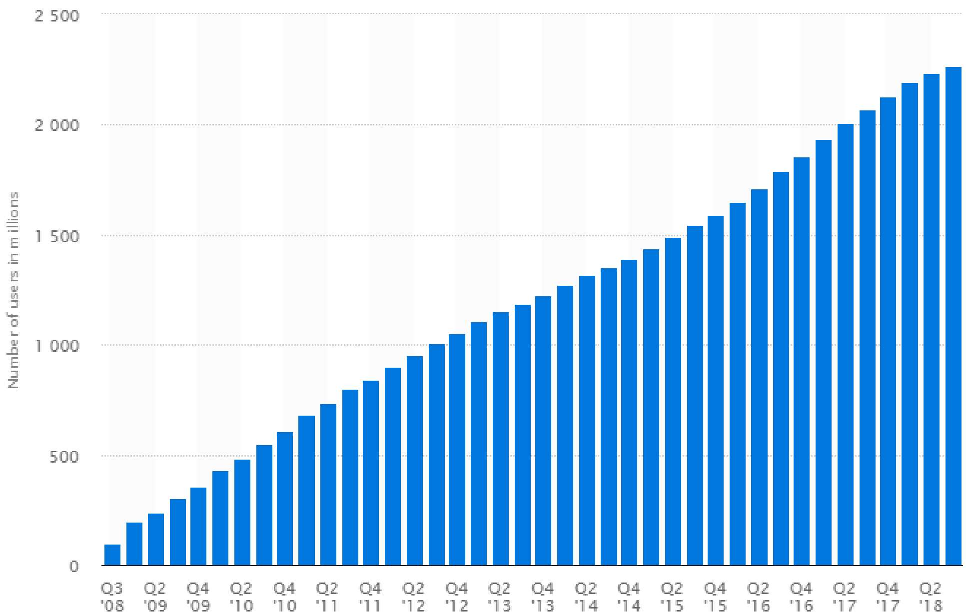
Even though MySpace monopolized the attention of media, SNSs start spreading all over the world. Friendster gained traction in the Pacific Islands, Orkut became the premier SNS in Brazil before growing rapidly in India [12] LunarStorm took off in Sweden, Dutch users embraced Hyves, Grono captured Poland, Hi5 was adopted in smaller countries in Latin America, South America, and Europe, and Bebo became very popular in the United Kingdom, New Zealand, and Australia. SNSs are now a global phenomenon.

### d. Social Media: Where we are now

In fact, we can say that Social Media Networks are a global phenomenon not only as a way of saying, but data allows us do say so.

The Social Media market is oligopolistic. In fact, we have few big sector leaders, of which one among the others can be called the king of Social Media, and we are talking about the groups owned by Mark Zuckerberg.

Started in 2004 as Harvard's student's platform, it took less than one year to reach 1 million users, less than 4 years to reach 100 million users and in 2009, after a little more than 5 years from its creation, Facebook became the most popular social platform in the world.

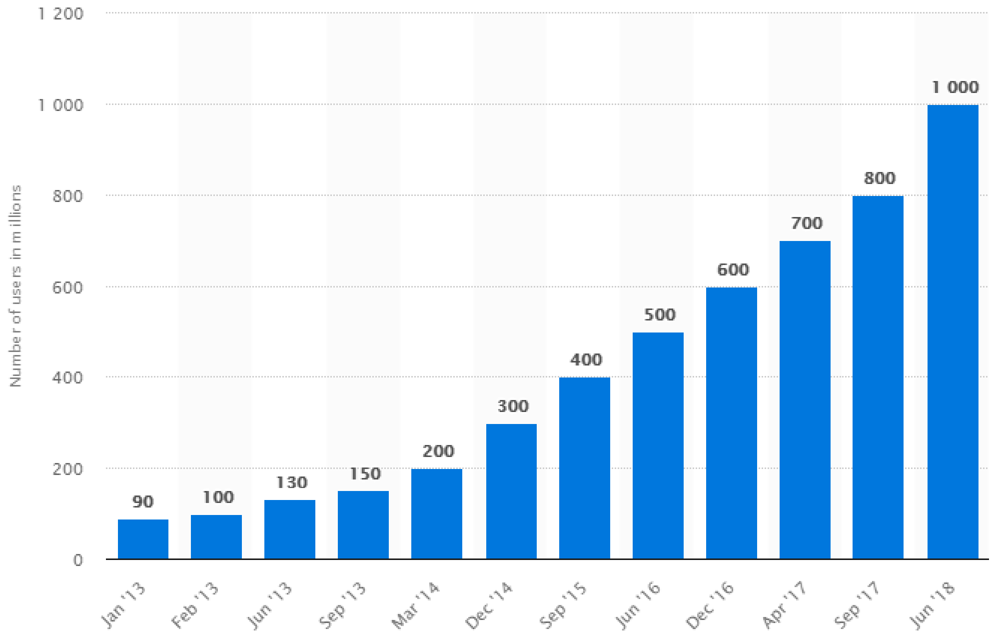



Tab A Facebook users from 2013 to 2018

In November 2010 Facebook was valued at a massive $41billion. Meanwhile, it became the 3rd largest web company in the US, sitting behind Google and Amazon.

Another huge milestone was reached the following year. June 2011 saw Facebook reach 1 trillion page views, according to a study by DoubleClick. And then, for the year overall, Nielsen found the site was the 2nd most visited in the United States.

In April we see Facebook make a major acquisition: Instagram. Spending $1bn, we get an idea of the kind of resources the platform now has at its disposal. This was just a month before the year's big event. The IPO.

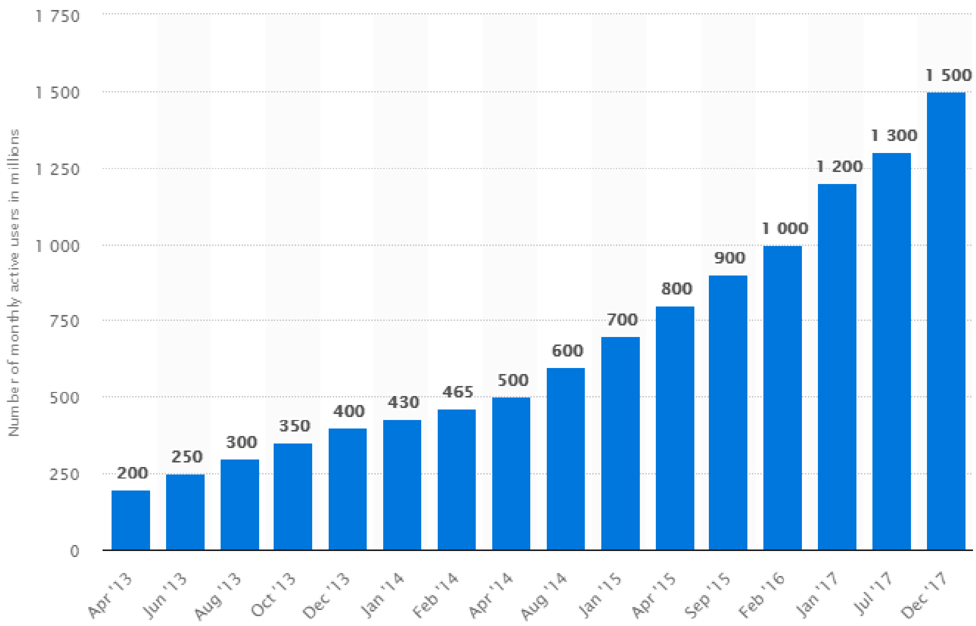



Tab B Instagram users from 2013 to 2018

It is easy now to make the math and see that the combined users from the two platform reach an outstanding 3.3 billion.

Nevertheless, Facebook didn’t stop here. In 2014 they reached 1 billion users log into the platform using a mobile device and in February of the same year, they announced they’d be buying WhatsApp for an incredible $19bn.

While Facebook already had a messaging system, this acquisition gave access to WhatsApps younger user base and their overseas users, reaching now 4.8 billion users from a wide age base.




Tab C Whatsapp users from 2013 to 2018

Out of 4.2 billion internet users, 3.03 billion are active users on Social Media, spending on average 116 minutes per day on Social Media. 91% of retail brands use 2 or more social media channels and 81% of all small and medium businesses use some kind of social platform.

These statistics are self-explanatory but, the ones referring to business data are even more astonishing.

In 2016, $40bn was spent on social network advertising in 2016 while 38% of organizations plan to spend more than 20% of their total advertising budgets on social media channels in 2015, up from 13% a year ago. [13]

These numbers cannot be ignored, they are the sign of a changing era inside the business world, a more connected, social and data-focused era.

# 3. Social Media Data Analysis

As what clearly emerged so far, numbers about social media are big.

They are not only big, but they are also easy to track, representative of the big and heterogeneous size of the population and always up to date.

When we talk about billions of users and trillions of data gathered from them, we must use a new term specific for this Brobdingnagian amount of bytes, we can talk about Big data.

Big data is a term used to refer to data sets that are too large or complex for traditional data-processing application software to adequately deal with. Current usage of the term "big data" tends to refer to the use of predictive analytics, user behavior analytics, or certain other advanced data analysis methods that extract value from data, and seldom to a particular size of data set.

"There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem." [14] Analysis of data sets can find new correlations to "spot business trends, prevent diseases, combat crime and so on."[15]

The Data Analysis is being also subject of criticism in the last period, with some of the biggest scandals like the most famous but for sure not the only one, Cambridge Analytica.

The Facebook–Cambridge Analytica data scandal was a major political scandal in early 2018 when it was revealed that Cambridge Analytica had harvested the personal data of millions of people's Facebook profiles without their consent and used it for political purposes, in particular for the 2015 and 2016 campaigns of United States politicians Donald Trump and Ted Cruz, 2016 United Kingdom Brexit vote and 2018 Mexican General Election [37].

We are paying for these services with data, and this fact exposes all of us, especially in our privacy, to threats we are not prepared for and that legislation doesn’t cover yet.

We live in a world of data: Wal-Mart, a retail giant, handles more than 1m customer transactions every hour, feeding databases estimated at more than 2.5 petabytes—the equivalent of 167 times the books in America's Library of Congress. [16].

Facebook, a social-networking website, is home to 40 billion photos and decoding the human genome which involves analyzing 3 billion base pairs—which took ten years the first time it was done, in 2003, but can now be achieved in one week.

All these examples tell the same story: that the world contains an unimaginably vast amount of digital information which is getting ever vaster ever more rapidly. This makes it possible to do many things that previously could not be done: spot business trends, prevent diseases, combat crime and so on. Managed well, the data can be used to unlock new sources of economic value, provide fresh insights into science and hold governments to account.

We are now creating so many data that, even with the latest technologies, cannot be analyzed and, actually, not even stored.

There are many reasons for the information explosion. The most obvious one is technology. As the capabilities of digital devices soar and prices plummet, sensors and gadgets are digitizing lots of information that was previously unavailable. And many more people have access to far more powerful tools.

For example, there are 4.6 billion mobile-phone subscriptions worldwide (though many people have more than one, so the world's 6.8 billion people are not quite as well supplied as these figures suggest), and 1 billion-2 billion people use the internet. The amount of digital information increases tenfold every five years.

Moore's law, which the computer industry now takes for granted, says that the processing power and storage capacity of computer chips double or their prices halve roughly every 18 months.

It is estimated that 80% of the world’s data is unstructured, but businesses are only able to gain visibility into a portion of that data. Innovative companies are using data to enhance their value proposition and increase customer satisfaction. [22]

We understood that data are an incredible source of benefits and, if not properly stored and analyzed, they are a heavy cost-opportunity for all the businesses that want to emerge and make the difference among the market.

The source of all these numbers is, in the end, people. Every time we surf on the internet, every time we search for a new receipt on the internet, we are generating data, and this data are getting a role in all the tech industries and the economic environment as well. What this paper is about, is a specific part of the data that we are creating every day, a still unexplored topic but with huge potential: the sentiment analysis.

# 4. Sentiment analysis

## a. Historical introduction

The object of sentiment analysis has typically been a product or a service whose review has been made public on the Internet.

This might explain why sentiment analysis and opinion mining are often used as synonyms, although, we think it is more accurate to view sentiments as emotionally loaded opinions.

The interest in other's opinion is probably almost as old as verbal communication itself.

Sentiment Analysis is one of the youngest and fastest growing research areas, with 99% of the papers appeared after 2004.

We observed that the first academic studies measuring public opinions are during and after WWII and their motivation is highly political in nature. The outbreak of modern sentiment analysis happened only in the mid-2000s, and it focused on the product reviews available on the Web.

Since then, the use of sentiment analysis has reached numerous other areas such as the prediction of financial markets and reactions to terrorist attacks

Additionally, research overlapping sentiment analysis and natural language processing has addressed many problems that contribute to the applicability of sentiment analysis such as irony detection and multi-lingual support.

## b. A general overview

Opinions are central to almost all human activities and are key influencers of our behaviors. Our beliefs and perceptions of reality, and the choices we make, are, to a considerable degree, conditioned upon how others see and evaluate the world.

Historically, leaders have been intrigued by the opinions of their subordinates to either prepare for opposition or to increase their popularity. Examples of trying to detect internal dissent can be found already at Ancient Greece's times.

Ancient works in East and West mingle with these subjects. "The Art of War" has a chapter on espionage that deals with spy recruiting and betrayal, while at the beginning of "Iliad" the leader of Greeks Agamemnon tries to gauge the fighting spirit of his men.

Voting as a method to measure public opinion on policy has its roots in the city-state of Athens in the 5th century BCE

"What other people think" has always been an important piece of information for most of us during the decision-making process, asking our friends to recommend a plumber, asking who they were planning to vote to end by checking consumers review on Amazon in order to decide what to buy.

Nowadays, thanks to the internet (among other things), it is possible to find out about the opinions and experiences of a vast pool of people, both professional critics or people who both it for a personal acquaintance.

We know that we search for others' opinion in order to make decisions, and data confirm these thoughts:

- 81% of Internet users (or 60% of Americans) have done online research on a product at least once;
- 20% (15% of all Americans) do so on a typical day;
- among readers of online reviews of restaurants, hotels, and various services (e.g., travel agencies or doctors), between 73% and 87% report that reviews had a significant influence on their purchase;
- consumers report being willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item (the variance stems from what type of item or service is considered);

- 32% have provided a rating on a product, service, or person via an online rating system, and 30%;
- (including 18% of online senior citizens) have posted an online comment or review regarding a product or service. [17], [18]

Not only people want to know what other people think about the products they are going to buy, but they are pushed by an insatiable impulse to write their opinions about the product they just bought.

Therefore there is no doubts about the vast amount of reviews that have been created every day influence the way we took decision on what to buy, meaning that for firms is always more critical to understand what people think about their product, if it is bad or good reviews, if they express happiness, sadness, if they love the product but it is too expansive, because the customer reviews are the first source of improvement.

Apart from real-life applications, many application-oriented research papers have also been published. For example, a sentiment model was proposed to predict sales performance [38] or again a model where reviews and 5-stars model were used to rank products and merchants. [39]

### c. Definition

"Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes". [20]

### d. Levels of analysis

The sentiment analysis has a huge potential due to the usefulness of the information that it offers but also thanks to its flexibility, making possible to analyze a bunch of different types of text or inputs, at different levels of depth, for different purposes.

The first classification of a vast topic as the above-mentioned one is the classification by depth of analysis that can be summarized into three different levels:

(1) Document level, this is the simplest and easiest level of sentiment analysis. The aim, in this case, is just to understand the positive or negative sentiment in a certain document, that can go from a paper, a product review or a short text.

(2) The second layer is the sentence level, where the task is focused on understanding if a single sentence is expressing a positive, negative or neutral opinion. The sentence sentiment analysis is becoming very popular as far as we are getting in depth inside the social media era.

Comments, tweets, short review, posts, everything nowadays is short and generated very frequently. It is mandatory to analyze all these sources of information in other to understand what people think about our product, the feeling expressed by the public towards a company or an institution or, in the end, the position of the electorate between two candidates running for the White House.

(3) The two levels just described don't analyze the opinion itself, but just if the document, text or paragraph is overall negative, positive or neutral. Although this can be useful to understand the feeling of the company's customers, these analyses don't answer to questions such as "What they like about the product?", "What they don't like?" "Is the price too high?".

Nevertheless, in case we want further insights about customers reviews or user's opinion, we can use the Aspect Sentiment analysis, the last one according to Bing [20]. The Aspect level works based on the definition of opinion, composed by a sentiment (either positive, negative or neutral), the target or better the thing that is being talked about and the opinion holder and the person expressing the opinion. Let's make an example to clarify the difference between the two kinds of analysis.

"This iPhone is amazing, but my mom thinks it is too expensive". The sentence-level sentiment analysis will just analyze, depending on how it is programmed, the sentiment of the sentence (most likely a product review on a website). It is effortless both for us and for a machine with the lexicon method, that we will analyze on a further chapter, that this is an overall positive sentence from the point of view of the writer.

This result though doesn't give us any insights on why it is positive and at which degree it is positive if the writer actually bought or will buy the product and which are negative aspects.

On the other hand, with the aspect sentiment analysis we are able to understand that: first, the sentence has a positive and a negative part. Furthermore, it tells us that there are two subjects in this review. The first subject is the iPhone itself and is reviewed positively by the writer since the positive adjective is referred to him.

However, the second part of the sentence tells us that the subject "my mom", bears a negative sentiment against the product, especially against one of its attributes: the price. With further analysis, simply analyzing the age of the writer, it would be clear that the opinion on the second part of the text is what would be in the end the decision making variable toward the purchasing or not of the product, with a series of considerations to be made, such as that to encourage the purchase of this target group it is worthless to work on marketing and advertisement on the children but better work on mums, lowering the price or better, put more efforts in showing the value.

However, it is worthless to say how, the second kind of analysis, is more complicated and trickier rather than the first one but, at the same time, gives richer information. [20]

### e. Types of sentiment analyses

Before going in depth in the topic, let's make a quick overview of the different types of sentiment analysis. Each one of them has a different complexity and gives a different grade of detail about information, so each one has a different scope.

*Fine-grained Sentiment Analysis*

This analysis goes a little bit more in depth than the standard sentence analysis. It is particularly useful when we need information about the overall sentiment of the sentence, but we also want the degree or the intensity of that sentiment.

The fine-grained analysis helps us with this by considering five different categories: very positive, positive, neutral, negative and very negative. Typically, these categories are expressed with the well-known 5-star rating system. Some of those systems can also provide the feeling associated with the sentiment (ager, happiness, sadness etc.).

*Emotion Detection*

It is natural that the main focus of this analysis is on the emotions expressed inside the opinion we are analyzing. Many of these systems are based on lexicon (a list of words divided in positive, negative and neutral) or complex machine learning algorithms.

Machines based on the lexicon are programmed following simple logic: they count the number of positive, negative and neutral words (following the rules and instructions given by the lexicon book or database) contained in a sentence. If the number of positive words is greater than the negative one it suggests a positive sentence, otherwise a negative one. Even though this methodology is fast and can be helpful, its' cons are easy to find.

First, it doesn't understand the context or the connotation is given to a certain word. For instance, the word "killing" (usually associated with a negative feeling) can be used in both a positive or negative way. "These prices are killing me" has a bad connotation; "You are killing it", on the other hand, has a positive feeling. The other con is that it doesn't consider the weight on each word.

Let consider the following sentence and read it with a lexicon-based machine point of view: “Even though this phone has a good display, a nice design, and a really good RAM, it doesn’t convince me into buying it” There are 3 times more positive words rather than negative ones, however, the sentence is indeed negative. [21].

*Aspect-based sentiment analysis*

This analysis is based on considering an object as a set of attributes. It focuses on the first time, on understating the sentiment of a sentence but, going further, also understanding to which of the characteristics or attributes of the product the opinion is related to. This analysis is perfect to analyze product reviews. "The battery life of this camera is too short." The sentence is expressing a negative opinion about the camera, but more precisely, about the battery life, which is a particular feature of the camera.

The increased utility of this variation of Sentiment Analysis is followed by a parietal increasing in the complexity for the machine to run the analysis, suggesting that it must be used for specific cases, narrowing down as much as possible the variability of the data given as input.

*Intent analysis*

Let’s analyze the three following sentences:

1. “Your customer support is terrible, after 20 minutes I didn’t solve my problem”;
2. “I would like to know how to reset my phone”
3. “Can you help me fill out this form?”.

A human being has no problems in detecting the complaint on the first sentence, the question on the second and the request on the third one, but it is a complex problem for a machine. It crucial for companies to invest and develop this kind of analysis because is a way to prioritize the request expressed by customers.

If a customer expresses a complaint should be served after someone expressing a request, for sure more urgent, as well as we can differentiate the number of question and the number of complaints, in order to understand if our customer service actually solve problems or if people are forced to ask again on our social media pages, website or any other channel.

*Multilingual sentiment analysis*

Multilingual sentiment analysis can be a difficult task. Usually, a lot of preprocessing is needed, and that preprocessing makes use of a number of resources. Most of these resources are available online (e.g. sentiment lexicons), but many others have to be created (e.g. translated corpora or noise detection algorithms). The use of the resources available requires a lot of coding experience and can take longer to implement.

### f. Sentiment Analysis algorithms

While choosing the machine that suits our requirements, according to the budget we want to allocate to the research, the time we have to run the machine and the depth of the results we want, we also have to take into consideration the type of algorithm used by the software.

*Concept-based approaches*

In particular, if we are talking about product reviews, the concept is more important then the sentiment expressed by the reviews, since the customer usually gives reviews about the firm or of the seller instead of the product, makes a comparison with other product from the same firm or competitors'ones.

Therefore was introduced the concept-level sentiment analysis, which focuses on a semantic analysis of text through the use of web ontologies or semantic networks, which allow the aggregation of conceptual and affective information associated with natural language opinions [45].

Unlike purely syntactical techniques, concept-based approaches are able to detect sentiments that are expressed in a subtle manner through the analysis of concepts that do not explicitly convey any emotion, but which implicitly liked to other concepts that do so.

The bag-of-concepts model can represent semantics associated with natural language much better than bags-of-words. In the bag-of-words model, in fact, a concept such as *cloud computing* would be split into two separate words, disrupting the semantics of the input sentence (in which, for example, the word cloud could wrongly activate concepts related to weather).

Also, without concept, product reviews can give fuorviant results once analyzed. Take the review *go read the book*: it has a positive polarity referred to a book review and a negative polarity referred to a movie review.

Also, the adjectives may be misinterpreted. The adjective *small* has a good polarity if referred to the queue at the post office, but a bad one if referred to a room in a hotel review.

*Rule-based approaches*

Usually, rule-based approaches define a set of rules in some kind of scripting language that identify subjectivity, polarity, or the subject of one opinion.

The most common algorithm in this category is the lexicon-based machine. It is the easiest to program with a set of rules that can be summarized in checking the number of positive and negative words inside a document, counting the two variants and return the category with more appearances.

Other machines can be based on Natural Language Processing techniques such as:

- Stemming,

This method allows reducing the vocabulary size by collapsing distinct word forms allowing us to reduce the run-time of our machine and sharpening the results given as outputs. There are three main algorithms in the context of sentiment analysis.

(1) Porter Stemmer, one of the earliest and best-known stemming algorithms, which works heuristically to identify word suffixes and strip them of, regulating the ending. Nevertheless, it increases the speediness of the analysis process, it can mislead the results because sometimes collapse and joint a positive and negative word into a single word, as you can see from Tab A.

| Positiv | Negativ | Stemmed |
|---|---|---|
| captivation | captive | captiv |
| common | commoner | common |
| defend | defendant | defend |
| defense | defensive | defens |
| dependability | dependent | depend |
| dependable | dependent | depend |

Tab A, word Positive/Ngative distinction destroyed by Porter Stemmer

(2) Lancaster Stemmer is based on the same principle on the Porter one, but for sentiment analysis can become even more problematic, collapsing more word and, even though it gives good results in terms of speediness and efficiency of the analysis, the results will be biased most of the time, make this algorithm not reliable in case we need a high accuracy rate.

(3) The most precise of the three is the WordNet Stemmer, with a few words identification destroyed in the stemming process. The results given using this algorithm are unbiased but, at the same time, the benefits in term of speediness and efficiency are sometimes not enough brilliant to justify the time spent to run the algorithm. The WordNet Stemmer collapse just tense, aspect and number marking, giving as output a vocabulary not collapsed enough. Consequently, although the high-precision of the results, it is not very widespread in the sentiment analysis context. [23]

- Tokenization

Tokenization is the process of breaking a stream of textual content up into words, terms, symbols, or some other meaningful elements called tokens.

Generally, the process of tokenization occurs at the word level, but it is sometimes tough to define what is meant by a "word".

Regularly a tokenizer commits on simple heuristics, for example, punctuation and whitespace may or may not be included in the resulting list of tokens, all contiguous strings of alphabetic characters are part of one token, in the same way with numbers and tokens are separated by the way of whitespace characters, such as a space or line break or by punctuation characters. The main use of tokenization is to identify the meaningful keywords, while the disadvantage is the difficulty to tokenize a document without any whitespace, special characters or other marks [24].

The most common Tokenization tools are Nlpdotnet Tokenizer, Mila Tokenizer NLTK Word Tokenize, TextBlob Word Tokenize, MBSP Word Tokenize, Pattern Word Tokenize, Word Tokenization with Python NLTK.

- Part of Speech (POS) Tagging

Even if POS tagging is not mandatory to run a sentiment analysis, there are many instances in which your system will incur in problems that POS tagging will solve. This technique allows to disambiguate different words with the same form, takes all the words in a text and tags them with its corresponding POS (its grammatical category, like “noun”, “verb”, “adjective”, “preposition” and so on)

For instance, by applying the POS tagging process, the verb “like” would become distinguishable from the preposition “like” [25].

After analyzing various types of rule-based algorithms, let go into practical use. A machine based on this rule has a really simple thus very long work. We must input one or more lists of words (Lexicon), correctly polarized, in other words with a tag for each word that says if the word is positive, negative, or neutral.

Then, we give to the machine the text to analyze, so the list of product reviews, tweets, articles or whatsoever. The machine will count all negative and positive words.

If the number of positive word appearances is greater than the number of negative word appearances return a positive sentiment, conversely, return a negative sentiment. Otherwise, return neutral.

This system is very naïve since it doesn't take into account how words are combined in a sequence. More advanced processing can be made, but these systems get very complex quickly. They can be very hard to maintain as new rules may be needed to add support for new expressions and vocabulary.

Furthermore, these approaches have many difficulties in pondering the words used in a sentence. Let’s make an example and try to read the following sentence by a rule-based machine point of view: “This product is nice, cheap, with a cool design and an average price/quality ratio, but I prefer the competitor’s product”.

In this sentence a rule-based machine would count the numbers of positive words and compare them with the negative ones, giving the overall results as positive, even though for every human being it is easy to understand that the sentence is negative.

*Automatic Approaches*

Automatic methods, contrary to rule-based systems, don't rely on manually crafted rules, but on machine learning techniques. The sentiment

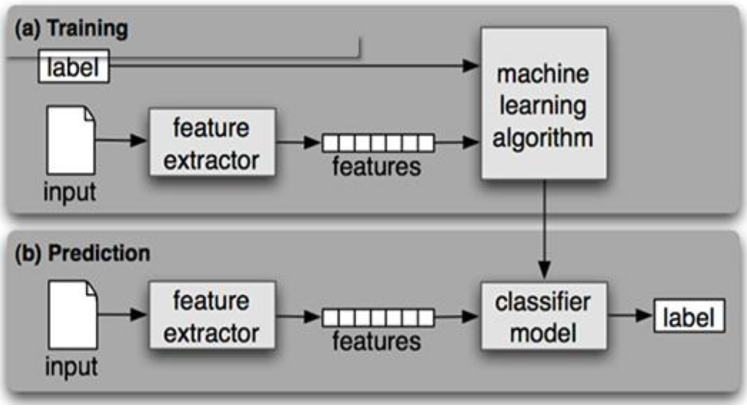



Fig. A, machine learning procss

analysis task is usually modeled as a classification problem where a classifier is fed with text and returns the corresponding category, e.g. positive, negative, or neutral (in case a polarity analysis is being performed).

A learning machine usually works following the steps and components showed if Fig A:

The overall process is easy to understand.

On the first time, the machine is fed with a model in order to associate a particular input with an output we choose, a tag o whatever on a test used for training. After that, on the prediction process, the machine is fed by an unseen text that it must recognize as positive, negative or neutral.

An automatic machine can also be based on classification algorithms, here at the main ones.

- Naïve Byes

The Naïve-Bayes algorithm is particularly widespread because of its simplicity and effectiveness, reliability and accuracy and for its success among the Natural Language Processing problems.

With this algorithm, belonging to the family of the probabilistic algorithms, the machine uses the probability theory and Bayes' Theorem to predict which tag should be associated with a piece of text. It calculates the probability of each tag of a given set and selects the one which has the highest probability. The algorithm used by these machines is the Multinomial Naïve Bayes.

In order to work properly, the machine governed by this algorithm needs to use a learning model with the correct features. Feature is a term which describes the piece of information we take from the text to give to the machine. For example, if we were doing classification on health, some features could be a person's height, weight, gender, and so on. We would exclude things that maybe are known but aren't useful to the model, like a person's name or favorite color.

Once we defined the features, the work is as simple as effective. We just use the word frequencies to define the possible outcome.

Nevertheless, we must acknowledge also the naïve part of this approach. We assume that every word in each sentence is independent of the other words and from the other sentences, ignoring the context and the person who wrote the text, ignoring the order of the words as well inside the text. By saying this, we imply that "I like the party" party I the like" and "The like party I" would be treated in the same way from the machine.

- Linear Regression

We are usually habituated to think about machine learning and sentiment analysis associated with something nebulous and complex, but there are also easy but effective algorithms applicable to the automatic machine. One of those is the Linear regression method.

This method uses the well-known algorithms in statistics used to predict a certain level of Y given a set of values of X. It is particularly easy to understand and apply, the only things required to the machine is the set of X values of our choice and the machine, applying the linear regression formula

$$Y = b_0 + b_1 * X$$

The machine will return the number coherent with the rest of the data given as input. Even though the algorithm seems easy, there are two situations in which we can find wrong outputs from the machine [26].

The first one is the underfitting, it means that the machine didn't learn enough from the training data given at the beginning, consequently, it will give results with an error too big to neglect.

On the other hand, if the machine overlearns from the training data, the output would be too sensitive, and the results would be useless. This phenomenon is called overfitting. The two phenomena are described in the graphs below [27].

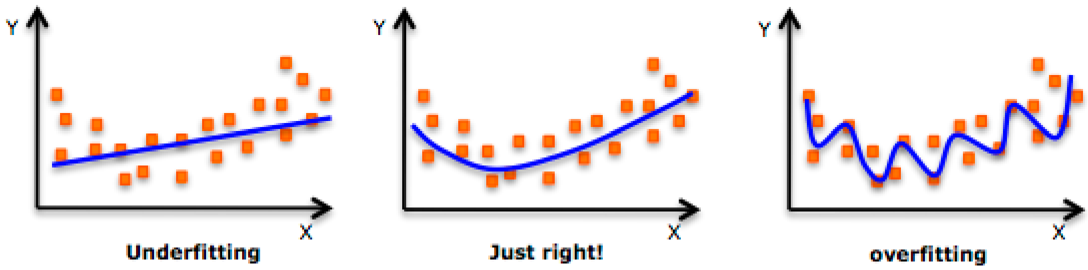



Fig B, difference of the results in case of underfitting or overfitting.

- Support Vector Machines (SVM)

This algorithm is considered as the next step after trying the Naïve Bayes one. It is well-known because of its great performance even in a situation of a very limited amount of input data.

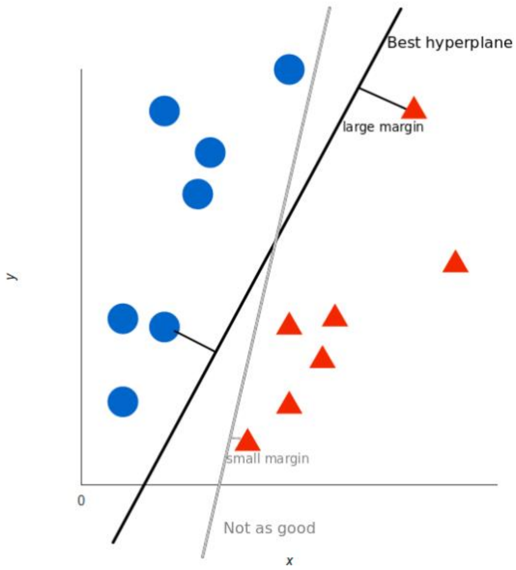



Fig C How is traced a Hyperplane

A support vector machine takes these data points and outputs the hyperplane (which in two dimensions it's simply a line) that best separates the tags. This line is the decision boundary: anything that falls to one side of it we will classify as feature A or Tag A, and anything that falls to the other as the other tag.

The hyperplane, for a Support Vector Machine algorithm, is the equation that maximizes the distance from both tags, as we can see in Fig. C. In other words, the hyperplane has the distances between the two nearer elements as largest as possible.

- Deep Learning

Last but not least, an automatic machine can be programmed based on a deep learning algorithm.

Deep learning is the application of artificial neural networks (neural networks for short) to learning tasks using networks of multiple layers. It can exploit much more learning (representation) power of neural networks, which once were deemed to be practical only with one or two layers and a small amount of data.

Inspired by the structure of the biological brain, neural networks consist of a large number of information processing units (called neurons) organized in layers which work in unison. It can learn to perform tasks (e.g., classification) by adjusting the connection weights between neurons, resembling the learning process of a biological brain.

Many deep learning models in NPL need word embedding as input features, it means that we need to use, before applying the deep learning model to our analysis, a software or a tool which is able to convert a vocabulary into vectors or continuous real numbers [28].

*Hybrid approaches*

The concept of hybrid methods is very intuitive: just combine the best of both worlds, the rule-based and the automatic ones. Usually, by combining both approaches, the methods can improve accuracy and precision.

# 5. Evaluating the performances of a Sentiment Analysis Machine

There are many ways in which you can obtain performance metrics for evaluating a classifier and to understand how accurate a sentiment analysis model is. This aspect of the analysis is important as much as the analysis itself. It is worthless to gather and elaborate Brobdingnagian amount of data, using the most powerful machine available in the market, if then we can’t measure the level of reliability of the results given by this machine. Thereupon, is mandatory, while dealing with these software, to understand how to evaluate their performance properly.

One of the most frequently used is known as *cross-validation.* What cross-validation does is splitting the training data into a certain number of training folds (with 75% of the training data) and at the same number of testing folds (with 25% of the training data), use the training folds to train the classifier, and test it against the testing folds to obtain performance metrics. The process is repeated multiple times and an average for each of the metrics is calculated.

If the results given by the cross-validation, we are facing a case of overfitting, as mentioned above, where the machine was trained too much against the same data set that, facing real data, the machine might fail to analyze properly a completely different set of data.

Furthermore, at a global level are informally defined a set of parameters which define the overall performance of the machine, and in particular, they are divided into 3 different aspects of the performance: Precision, Recall, and Accuracy.

On the first hand, we should measure the *Precision.* This indicator measures the number of texts predicted correctly as belonging to a certain category out of all the text (correctly and incorrectly) predicted as belonging to that category. We can think about this indicator as an error ratio per category, giving the information of the percentage of text predicted correctly over all the text predicted in a certain category.

On the other hand, we have the *Recall* indicator. This index is measured by category as the previous one, and it is calculated with the ratio between the numbers of correct prediction of the sentences in a certain category divided by the total prediction for that category. The more information about the abovementioned category we gave as inputs to the machine, the more accurate will be the precision ratio.

Last but not least, we should always take into consideration another important key element for a sentiment analysis machine: *Accuracy*. It may seem similar to the previous indicator, but the accuracy of a machine is measured within all the text, not just a category. It is calculated by dividing all the texts predicted correctly out of all the texts given as outputs from the machine.

# 6. Main Challenges for a sentiment analysis machine

There are a lot of different machines, with different complexities but they all have to face the same kind of problems. The first, most general, and source of all the other problems is the intrinsic complexity of the human written expression.

Doesn't matter how long and detailed is a Lexicon, it will never be able to forecast every word expressed by a human being and even less forecast the combination of words used by him, considering dialects, a way of saying, grammar or logical mistakes and so on.

So far, we can say that the task of reading what millions of people think is complicated. Nevertheless, one of Moore's Law's corollary express that, a certain point in human history, if not even now, we will have enough computing power to crack also this challenge.

*Subjectivity and Tone*

How many times happen to think "What did she mean with that sentence?".

For a sentiment analysis machine is the same, but way more complex since it has no human brain to understand the tone of the sentence and, by changing the tone, a sentence can be the opposite.

Secondly, the subjectivity analysis is crucial as well as the tone analysis. In fact, objective tests do not contain sentiments, consequently, they should be all categorized as neutral.

Taking into consideration the following sentences "The package is nice" and "The package is red". Most people would say that sentiment is positive for the first one and neutral for the second one, right? All predicates (adjectives, verbs, and some nouns) should not be treated the same with respect to how they create the sentiment. In the examples above, nice is more subjective than red.

Nevertheless, the complexity does not end here. Some adjectives can be objective in some context, (for instance in “the car is red” we can say without doubts that is objective), but they can have a big subjective connotate in another context “My mobile phone screen was all red”, underlining an impatience and anger sentiment.

*Context and Polarity*

All utterances are uttered at some point in time, in some place, by and to some people, you get the point. All utterances are uttered in context. Analyzing sentiment without context gets pretty difficult. However, machines cannot learn about contexts if they are not mentioned explicitly.

There are no easier ways to explain this concept if not with an example. Let’s take into consideration these two tweets, “Everything of it” “Absolutely nothing”.

Imagine how much can the polarity change by shifting the context where these two sentences are referred to. If the question was “What did you like about this product?”, then the first one is extremely positive and the second one the contrary, but if the question was simply asked with a slightly different connotation like “What did you DISlike about the product?”

Of course, it changes everything, and communicate the context to the machine is a process which requires an incredible amount of processing power both in the pre and post processing phase.

However, how to pre-process or post-process data in order to capture the bits of context that will help analyze sentiment is not straightforward.

*Irony and sarcasm*

Understanding irony and sarcasm is a tough task even for a normal human being, requiring a deep understanding of human language and a big dose of intelligence.

According to the Cambridge dictionary, irony is “a situation in which something which was intended to have a particular result has the opposite or a very different result” and, the most insulting or ridiculizing version of irony (also known as sarcasm) is “the use of remarks that clearly mean the opposite of what they say, made in order to hurt someone's feelings or to criticize something in a humorous way”.

These two language properties usually change positive sentiment into negative whereas negative or neutral sentiment might be changed to positive. However, detecting irony or sarcasm takes a good deal of analysis of the context in which the texts are produced and, therefore, are really difficult to detect automatically.

As always, a clever way to explain how a machine works compared to a human mind is to force the human mind first to think, so here the example. Let’s take into consideration these two sentences: “Yeah. Sure.” “Not one, but many!”. If we are familiar with the English language, we will not have any difficulty in finding the sarcasm in the first sentence, probably because we head that example use with sarcasm a lot of time during our life.

The problem is there is no textual cue that will make a machine learn that negative sentiment since most often, yeah and sure belong to positive or, eventually, neutral texts, and teach to machine how to detect it is a real challenge.

How about the second response? In this context, the sentiment is positive, but we’re sure you can come up with many different contexts in which the same response can express the negative sentiment. It is enough to give some context in order to make the sentence negative, imagine if the answer is to the question

"Did you encounter problems in dealing with our customer service?". The second example is for sure negative.

But to make it more difficult, think about that answer to a question like: "Did you try one of our products?". Without further information, it is almost impossible to even for a human being to understand if there is sarcasm in the sentence.

*Comparison*

Another challenge in the complex word of sentiment analysis is the comparison, but it is a challenge worth to tackle.

Look at the following texts: "This product is second to none." "This is better than old tools." "This is better than nothing."

There are some texts that don't even need a context to be correctly analyzed. Even a machine without a lot of training data would return the first sentence as positive, and for sure a normal person would not a problem as well.

However, if we take a closer look at the second and the third sentence, we start encountering some difficulty even in giving the simple classification as positive, negative or neutral.

Starting from the second one, we cannot say if what the customer meant is positive or neutral if we don't know anything about the old tools. If they are good tools, then it became a very positive sentence, but if they are not good the text becomes neutral or if the increase is too little, even negative.

Once again, context makes a huge difference.

Things become even more complex tackling the third sentence. Even a human being here would have difficulties in labeling the sentence, challenge that would be just slightly facilitated.

*Emojis*

We are almost entering the third decade of the XXI century, and we cannot talk about modern communication language without taking into consideration the emojis.

Studies on Twitter are becoming quite common these years. Even so, the majority of them did not focus on emoticons, even less on emojis. Nevertheless, there will be no one under the 20 years old that would say to not use them, making the emojis integrant and fundamental part of the digital language spoken nowadays by both Millennials and Z Generation.

Quoting [33], there are two types of emojis, even though the distinction is rather unofficial (meaning categorized by users) but old enough, categorized in western emojis and eastern emojis. This distinction is important as most of sentiment analysis works do not consider eastern emoticons but only western emoticons.

The term emoticon is a shortcut for "emotion icon", and we can say that western emoticons are the most commonly used and known emoticons, they usually are horizontal and have limited representativeness (=), xD, etc ).

On the other hand, eastern emoticons also known as "Japanese emoticons" or kaomoji, are vertical and can represent more complex faces and body positions than western emoticons. Such as o(0 0)/ (surprised with an arm up), /(Q.Q)\(crying, sad with arms downside).

Talking about the challenges these apparently simple language forms add to a sentiment analysis machine, we must start by the sentiment expression. A sentence with a neutral emotional state can be completely changed by the use of a happy, neutral or sad emoji.

Furthermore, the emojis can change the sentiment enhancement, allowing the writer to disambiguate an emotional sentence by pointing at a specific emotion when several may be present.

Most important in the context of the sentiment analysis is the faculty of an emojis to modify completely the sentiment of a particular text, allowing the reader to understand the eventual irony or sarcasm, adding information without using a single character.

It is pointless at this point to add how subjective are the emojis and how difficult it is to teach to a machine id they are used to modify, enhance or strengthen a sentiment into a text or, eventually, just for fun. [33]

Usually, sentiment analysis machines face the emojis challenge following different approaches.

- Keyword spotting approach for emoticons

Keyword spotting is often applied with an external resource such as an emoticon dictionary or sentiment lexicon. Related works combined emoticon lexicons and word sentiment lexicons to improve the supervised sentiment classification. More recently it has also been developed software able to replace emojis with their associated polarity, reducing the mistakes made by machines by deleting the neutral ones.

- Machine Learning approaches for emoticons

In machine learning approaches emoticons are used as a feature among others. The way this feature is used or represented differ from research work to another. We can encounter *emoticons as verified sentiment labels,* where each emoticon is used to define the polarity without even considering the rest of the sentence (usually “:)” and “:(”). Moreover, we can find a machine based on the number of emoticons, with a mechanism similar to the previous feature, basing the polarity on the biggest between the number of positive and negative emojis.[34]

As far as we can say, emojis are easy to use, but not easy to analyze in term of the case we are studying.

*Defining Neutral*

Moving from the difficulties given by the intrinsic complexity of our language to the ones more technical, defining what we mean by neutral is another challenge to tackle in order to perform accurate sentiment analysis.

Defining the standard of neutrality will define the entire output of the machine. Neutrality unbalanced can give a polarity completely different from what is the real sentiment of the subjects in the analysis. Since tagging data requires that tagging criteria be consistent, a good definition of the problem is a must.

A good neutral tag must contain objective texts first. As explained before, an objective text does not contain explicit thoughts of the customer, consequently, it doesn't contain sentiments. All the texts which fall in this category should be classified as neutral.

Another neutral tag should be reserved for irrelevant information. It is necessary in this case a filter for the information given as input to our sentiment analysis machine in order to reduce the amount of work of the software and, doing so, reducing the mistakes and improving at the same time the accuracy. It is important to be extremely aware of how the overall performance of the machine would change with these kinds of filters, sometimes they just add noises to the classifier modifying the outputs and, eventually, getting the platform performances getting worse.

Another element to take into consideration while defining neutral tags is wished. Some wishes like I wish the product had more integrations are generally neutral. However, those including comparisons, like I wish the product were better are pretty difficult to categorize.

*Subjectivity detection and opinion identification*

Work in polarity classification often assumes the incoming documents to be opinionated. For many applications, though, we may need to decide whether a given document contains subjective information or not, or identify which portions of the document are subjective.

# 7. Measuring the accuracy of Sentiment Analysis

Since the complexity of the multiple linguistic categories, sometimes not even during the annotation is possible to define doubtless a certain category. In order to solve this problem, or better, to measure the accuracy of the prediction, we use the inter-annotator agreement.

According to Langacker, “Inter-annotator agreement is a measure of how well two (or more) annotators can make the same annotation decision for a certain category.” [29]

And from this measure, we can extrapolate two things:

(1) how easy the categories are to define, if the two or more annotators give the same results for almost all the categories, it means that the guidelines we gave as input to these machines are clear and, in addition, that the borders between categories are very well defined.

(2) How trustworthy is the annotation. If the inter-annotator agreement were low, the annotators found it difficult to agree on which items belong to the category, and which didn’t.

There are basically two ways of calculating the inter-annotator agreement. The first approach is nothing more than a percentage of overlapping choices between the annotators. This approach is somewhat biased because it might be sheer luck that there is a high overlap.

Indeed, this might be the case if there are only a very limited amount of category levels (only yes versus no, or so), so the chance of having the same annotation is a priori already 1 out of 2. Also, it might be possible that the majority of observations belongs to one of the levels of the category so that the a priori overlap is already potentially high. [29]

The second and most widespread metrics for the inter-annotator agreement is the Krippendorff’s Alpha.

According to some researchers, the best inner-annotator agreement on Twitter sentiment analysis has 0.655 of Krippendorff's Alpha. This means that the value of the agreement is good, since it is greater than zero, but it still far from the value that social scientists use as reliability prove, that should be at least 0.8.[49]

A 0.655 expresses the intrinsic difficulty of the sentiment analysis, for the machines as well as for human beings. Taking into consideration that machine learning algorithms need input data to start the learning process, we can just assume that the error given by the machines are just the reflex of the human disagreement embedded on the data, making sentiment analysis even more challenging.

Sentiment analysis is a tremendously difficult task even for a human being, an inter-annotator agreement a is pretty low and that machines learn from the data they are fed with. That said, we must admit that sentiment analysis classifiers might not be as precise as other types of classifiers. sentiment analysis classifiers might not be as precise as other types of classifiers.

Considering what we said so far, you might be saying, "is it worth the effort?" The answer is simple: it sure is worth it!

Chances are that sentiment analysis predictions will be wrong from time to time, but by using sentiment analysis you will get the opportunity to get it right about 70-80% of the times you submit your texts for classification.

For typical use cases, such as ticket routing, brand monitoring, and VOC analysis this means saving a lot of time and money which in the nowadays frenetic research for the maximum efficiency can change the destiny of a company.

# 8. Sentiment Analysis applications

Sentiment analysis can be applied in many different methods and, thanks to its flexibility and adaptability, the only limit is in the user fantasy and capacities. Here are the most common uses.

## a. Social Media and Brand Monitoring

Nowadays, all companies have social media accounts, from where they can gather a lot of direct information about what their customers like or dislike about their product, their company, their action and so on.

Nevertheless, the world wide web can be used also to gather information in a wider way, accessing blogs, forums, news etc. In other words, a company can access all the mention about itself in the term in volume but, even more, in terms of quality of the mentions.

The Sentiment Analysis can be used to analyze news articles, blog posts, forum discussions, and other texts on the internet over a period of time to see the sentiment of a particular audience. It can be used to automatically categorize the urgency of the mentions about the brand online.

By analyzing the topic, the machine can understand the area inside the company related to the problem and assign them to the proper member of the team. Last but not the least, all these processes can be automated, allowing us the gather a lot of precious insights that, properly examined, allows us to position our brand in the market and among the competitors.

The main advantages of implementing this type on analysis within a company are that it gives the brand positioning in the market not only in a single moment but also its evolution through a defined period of time.

By applying the same analysis to our competitors' brand, we can position ourselves among them, see the mistake they made or their weak points and improve our company in order to fill that market gaps.

Moreover, we can identify potential PR crises and know to take immediate action, as well as tune into a specific point in the time to understand what the public thinks about us during the launch of a new product or during the day of our IPO.

b. Customer Feedback

Social media and brand monitoring offer us immediate, unfiltered, invaluable information on customer sentiment. In a parallel vein run two other troves of insight –surveys and customer support interactions.

Teams often look at their Net Promoter Score (NPS), a tool to measure customer experience and predict business growth [50], but we can also apply these analyses to any type of survey or communication channel that yields textual customer feedback.

NPS surveys ask a few simple questions – namely, would you recommend this company, product, and/or service to a friend or family member? and why? –and use that to identify customers as promoters, passives, or detractors. The goal is to identify overall customer experience and find ways to elevate all customers to “promoter” level, where they theoretically will buy more, stay longer, and refer other customers.

On the other hand, Sentiment analysis is useful in understanding the Voice of Customer (VOC) because it helps you to design better-informed questions to ask on future surveys, understand the nuances of customer experience over time, along with why and how shifts are happening, empower internal teams by giving them a deeper view of the customer experience, by segment and by specific aspects of the business.

In Brazil, federal public spending rose by 156% from 2007 to 2015 while people’s satisfaction with public services steadily decreased. Unhappy with this counterproductive progress, the Urban-planning Department recruited McKinsey to help them work on a series of new projects that would focus first on user experience, or citizen journeys when delivering services. This citizen-centric style of governance has led to the rise of what we call Smart Cities.

McKinsey developed a tool called City Voices, which conducts citizen (customer) surveys across more than 150 different metrics, and then runs sentiment analysis to help leaders understand how constituents live and what they need, in order to better inform public policy. By using this tool, the Brazilian government was able to surface urgent needs –a safer bus system, for instance– and improve them first. [40]

Again, is well known for the use of sentiment analysis to adjust the aim after a bad marketing campaign from Expedia Canada. It all goes back to Christmas 2014, Expedia ran, as always, an “Escape Winter” campaign, launching a commercial with a man looking at the snow accompanied by a Hitchcock-style screeching violin.

Customers just hated the violin and they were loud about it. Thousands and thousands of reviews were written by the customers on Social Media, blogs and forums.

Nevertheless, Expedia avoided big image damage with a counter campaign with a commercial featuring the violin being thrown in the snow, smashed by one of the original actors or even inviting a customer to rip the violin away.[46]

Even though their original product was far from flawless, they were able to redeem themselves by incorporating real customer feedback into continued iterations thanks to sentiment analysis and machine learning.

### c. Customer Support

“The customer is the King”, “be sure to have an amazing customer experience”, “is all about the customer”. Particularly in recent years, there’s been a lot of talk around customer experience and customer journeys.

Leading companies have begun to realize that oftentimes how they deliver is just as (if not more) important as what they deliver. Nowadays, more than ever, customers expect their experience with companies to be immediate, intuitive, personal, and hassle-free. In fact, research shows that 25% of customers will switch to a competitor after just one negative interaction.[41]

To avoid this to happen, we can automate systems to run Sentiment Analysis on all incoming customer support queries and rapidly detect disgruntled customers, surfacing those requests to the top.

After understanding the sentiment and the topic of each request, they can be prioritized and sent to the person best suited to answer or, if we want general information, we can see the overall sentiment towards the customer support, underlying its weaknesses and giving so the tools to improve it.

### d. Workforce Analytics and Voice of the Employee

In the same way, we measure VoC via customer surveys, we can solicit and act on feedback from our own employees. Chances are they are wildly more invested in giving actionable ideas on how to improve the workplace.

After sending a survey to our workforce, we can use the analysis to track the overall sentiment over time and by segment, extracting keywords and surface urgent concerns.

To show the benefits of such use of the tool, we can use the simplest case possible but, unfortunately, also a very common one.

Let’s say we conduct an internal survey asking employees to rate various aspects of their workplace experience and explain why they feel that way. On a scale of 1 to 10, a top-performing employee may say she rates her engagement at work as a 5 –not ideal.

However, if we look closer, we see she added: “I love the work I do and my training opportunities have been excellent, but my boss makes occasional inappropriate remarks towards me that make me feel uncomfortable.”

A response like this should raise red flags of potential sexual harassment and be brought immediately to HR’s attention in order to address the situation. If you simply slop it together with the other aggregated scores and don’t read through them for another two months, you risk losing a valuable employee or heightening an already tense situation.

That is why, in the case of big companies where it is not possible to go through every survey personally, it is still mandatory to analyze every comment and every remark.

### e. Product Analytics

In our agile world, we’ve learned that products are best built by prototyping early, soliciting feedback frequently, and continuing to iterate and improve. There is a specific branch of Lean Thinking, called Lean Design, which focuses on developing a product with a lot of feedbacks and continuous adjustments.

But for many product teams, soliciting frequent feedback can be the trickiest part. How do you narrow down which customer segment to ask? How do you sort through and weigh all their feedback? This is exactly where sentiment analysis can change the game. Whether by analyzing surveys, customer support interactions, or social media, machine learning enables you to assess huge amounts of product feedback at once.

With Sentiment Analysis it is possible to analyze large quantities of product reviews at once, especially in the early stages of the product development cycle, allowing the company or the developing team to make the necessary adjustments early in the process and avoid heavy and costly later changes.

Additionally, you can keep constant tabs of what people like and don’t like about the product, addressing the relevant comments to the project team. It is possible to find preferences of the different customer targets for different product features, allowing the producer to remove the costly and useless features for targets who do not need them

f. Market Research and Analysis

And as a final use case, sentiment analysis empowers all kinds of market research and competitive analysis. Whether you’re exploring a new market, anticipating future trends, or keeping an edge on the competition, sentiment analysis can make all the difference.

The first use on this case is to compare product reviews with competitors’ ones, generate periodic reports to understand how sentiment and trends are evolving, compare trends across the international market to select which country is more suitable to be one on which expanding, analyze tweets and social media post for instantaneously feedbacks and real-time happenings.

In this case, the most useful result we would have in running a sentiment analysis is to quantify otherwise qualitative information, adding also a qualitative dimension to already-gathered quantitative insights.

# 9. Choose Twitter advertisement strategies

Online Social Networks (OSNs) such as Facebook, Twitter, and YouTube have emerged as highly engaging marketing and influence tools, increasingly used by advertisers to promote brand awareness and catalyze word-of-mouth marketing.

In this case study, we analyze which advertisement strategy is more impactful depending also on the outcomes expected from the company. The analysis is performed both in terms of volume and sentiment analyzing the two main advertisement strategies on Twitter: *promoted tweets* and *promoted trends.*

Starting from the basis, a promoted tweet is a regular tweet that appears on the twitter feed without disappearing, as a normal tweet does after more or less 18 minutes [47].

The aim of a company, in this case, is to advertise the company itself and especially one or more products. The reactions expected by the public are likes and retweets.

On the other hand, promoting a trend is similar but we don’t want to promote a single tweet or a single product, we want to promote and hashtag to make it viral and attract more people to our page and, therefore, to our company.

Promoting a trend is way more expensive, on average $ 200 000 per single trend [48], because a company acquires the post with the trend, which remains in the feed as much as a promoted tweet, but the hashtag also appears at the top of the Trending Topics list on Twitter, and are labeled as a promotion.

The aim, with the promoted trend, is of course to reach the highest number of likes and retweet on the original post, but mainly to have as many people as possible to use the hashtag.

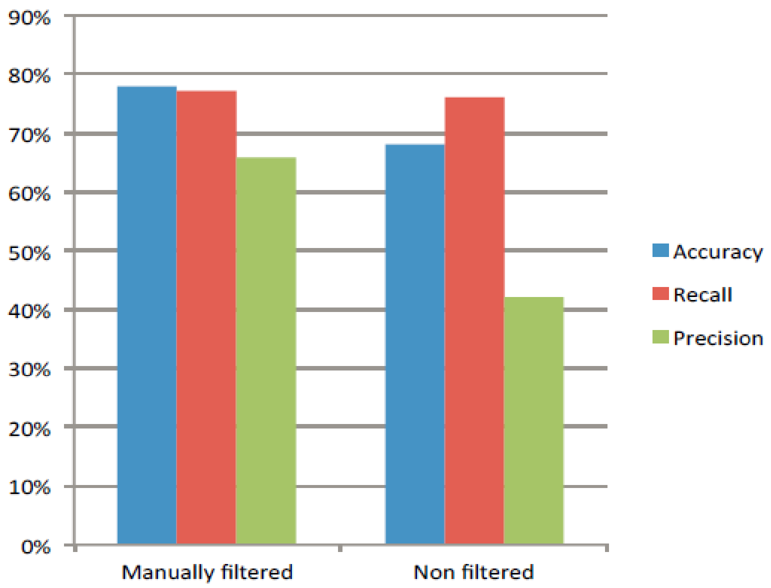
In this analysis, we enlighted a problem for sure relevant also in the other case studies, even though will not be touched again in order to avoid redundancies: the cleaning of the dataset.

One of the major challenges during cleaning the dataset and removing spam is ensuring topic relevance. However, during the cleaning, an issue must always be taken into consideration: A keyword-based approach tends to be too broad to accurately identify tweets referring to a particular brand.

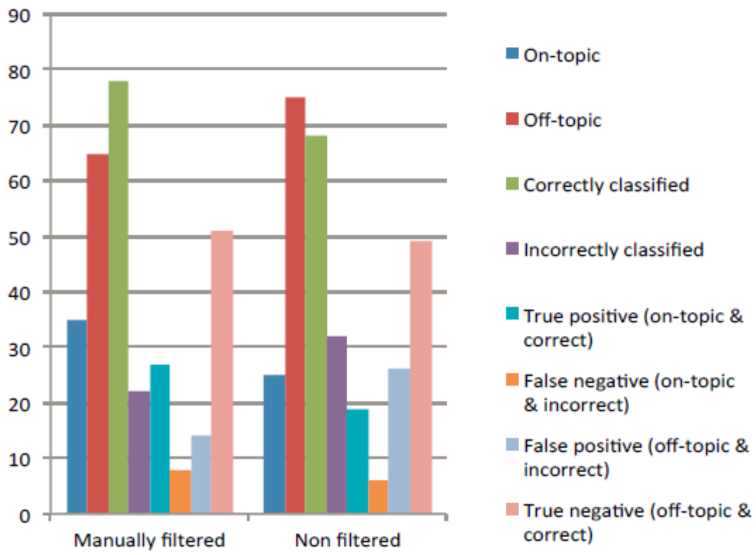
For instance, taking the UK mobile telecommunications provider and network $O_2$, research shows that over a 10 days period and a 90 000 tweets keyword analysis, just the 30% of the tweets were actually referring to the company taken into consideration, the others were about other companies or the Oxygen.

The problem can be solved using different classifiers. There is no best one, every analysis requires a different machine learning algorithm. To test each algorithm you need to manually label a certain number of tweets, depending on the size the population of tweets we take to analyze (on this specific case 200) and give them as input to our machine.

For this case, the best algorithm is Naïve-Bayes, with an accuracy of 84%.




Tab A: Classication results using Naïve Bayes.



Tab B: Classication details per class using Naive Bayes.

Looking at Tab A and Tab B we can see the benefits of training the machine even with a small size of tweets.

The version with manual filtering achieves 78% accuracy, 77% recall, and 66% precision. Tab B gives details of the per-class predictions: without manual filtering, false positives are more common than false negatives, meaning that too much irrelevant data is slipping through. Moreover, levels are much closer to filtering.

After cleaning the dataset, another important task is to define the method for the analysis. For a more robust tool for comparison, we examined two alternatives. As a hybrid lexicon/machine learning tool we chose SentiStrength.

This method uses a predetermined list of words commonly associated with negative or positive sentiment, which are given an empirically determined weight.

However, even though their word lists and weightings are determined for Online Social Network data (including Twitter), this approach may suffer when faced with social text with new words, unexpected spellings, and context-dependent language and meaning.

For a purely ML-based option, in this case, the study is used Chatterbox's Sentimental API, based on statistical machine learning over large, distantly labeled datasets. This data-based approach means it might be expected to handle slang, errorful or abbreviated text better.

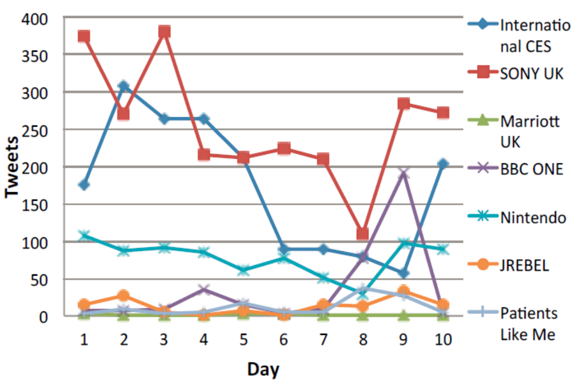
Before starting the analysis, a few hundred tweets were extracted as a sample and manually labeled, in order to test the accuracy of the machine, showing an accuracy of 84%.

After the "Analysis Set-up" is completed, a sample of tweets from 12 different companies in 7 different sectors was gathered, over a period of 10 days. Of course, just the tweets in English were taken into consideration and also the type of promotion was considered. Tab C shows the subjects of the analysis.

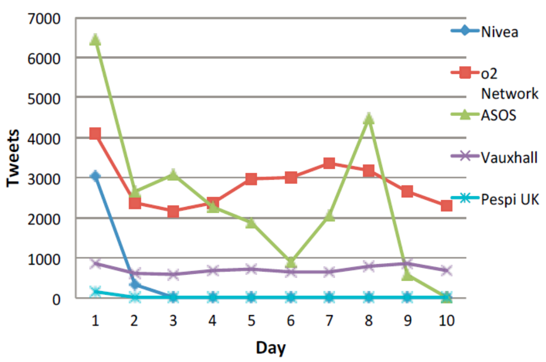
| **Industry** | **Promotion Type** | **Brand** |
|---|---|---|
| Electronics | Promoted tweet<br>Promoted tweet<br>*Promoted trend* | International CES<br>SONY<br>Nintendo |
| Travel | Promoted tweet | Marriot |
| Entertainment | Promoted tweet | BBC One |
| Automobile | *Promoted trend* | Vauxhall |
| Health Care | Promoted tweets<br>*Promoted trend* | Patients like Me<br>Nivea |
| Retail | *Promoted trend*<br>*Promoted trend*<br>Promoted tweet | ASOS<br>PepsiMax<br>Jrebel |
| Telecommunications | *Promoted trend* | O2 Network |

Tab C: Industry sectors and sample brands

To better analyze the output of this case study, to figures are taken into consideration: the volume of the tweets and their sentiment.




Tab D: Distribution of promoted tweets volumes over time.




Tab E: Distribution of promoted trends volumes over time.

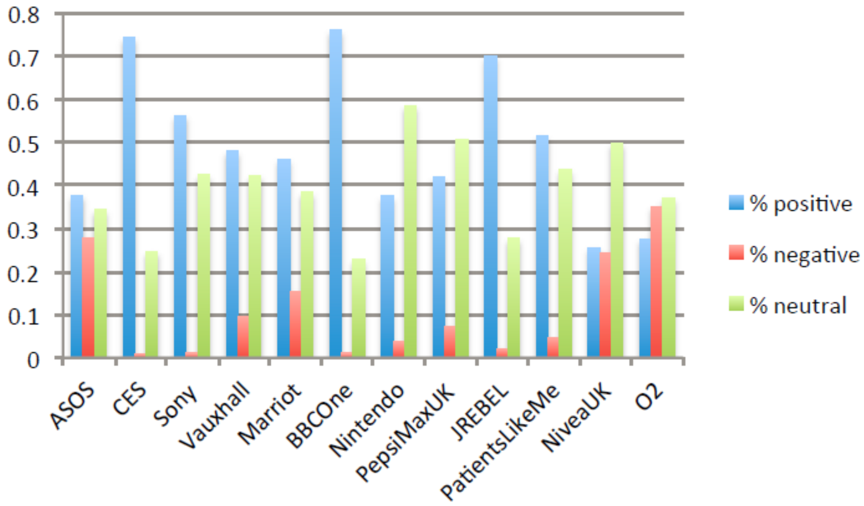
By a quick look to Tab D and Tab E, we can see that we are speaking about different orders of magnitude.

On average, promoted trends led to much higher response volumes. However, the highest percentages of mentions within responses were from promoted tweets, where an average of 18% of tweets each day included an `@' mention to the brand, against a 15% for promoted trends.

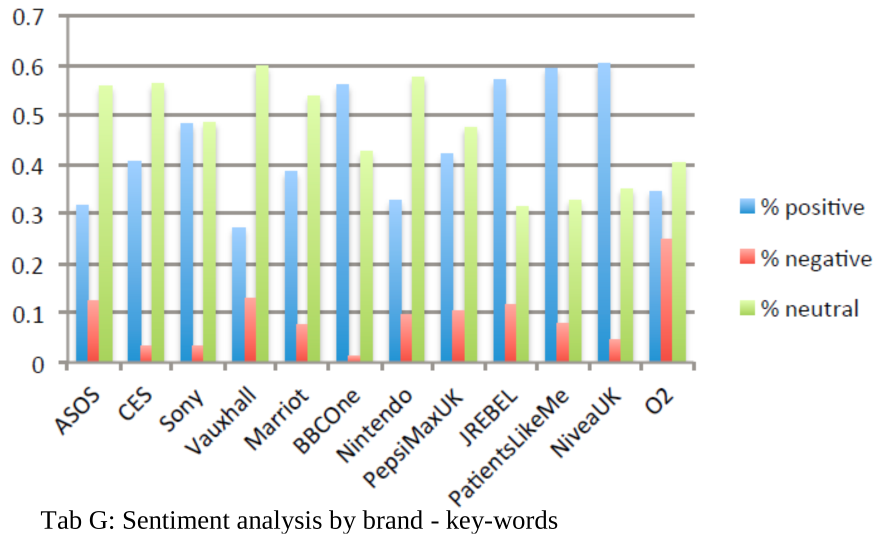
Moreover, the data shows that the greatest percentage of engagement for a brand's promoted item takes place on the first day of promotion reaching, on average, 24% of the total engagement with a steady decline over time, reaching a certain level of engagement with rises and falls, never reaching the first day though.

The effect is most pronounced for promoted trends, with 34% of engagement on average on the first day of the promotion, after which the engagement falls dramatically by an average of 25% to 9% by day two and continues to fall thereafter, even if the item is promoted for several days.

In general, though, these results show that adoption of a promoted item is not a slow gradual shift over several days (as might be assumed) but rather an immediate incline when exposure to the item is new to users.




Tab F: Sentiment analysis by brand - machine learning




Tab G: Sentiment analysis by brand - key-words

On average, across all brands (promoted tweets and trends), the average percentage of tweets and retweets12 which contained a positive sentiment is 50%, that which contained a negative sentiment is 12%, and 38% of tweets had a neutral tone.

# 10. Trump vs Clinton

The case I propose is the analysis of Twitter in order to anticipate and predict the winners of the election. Such a difficult task, considering we have just 140 characters to write on this media. But let’s start from the basics. Elections are not anymore as they were in the past, now the online campaign is as much important as the offline one, if not more important.

Everything started in 2008, with the first presidential campaign of Barak Obama. The successful use of social media in that campaign has established Twitter, Facebook, Myspace, and other social media as integral parts of the political campaign toolbox. Some analysts attribute Obama's victory to a large extent to his online strategy.

A study conducted by A. Tumasjan and T. Sprenger [35] shows that Twitter can be actually be used to predict the social sentiment towards candidates, analyzing the French and German elections between 2007 and 2009.

What I will bring instead, is the sentiment analysis of something more recent, and, as far as I’m concerned, more impactful: the US presidential election of 2016, with the race between Hillary Clinton for the Democratic Party and Donald Trump for the Republican Party.

The American election is not anymore a national matter. During the campaign, all the world looks at the States and comment on the social media what it thinks. To analyze the campaign millions of tweets were collected, analyzed and polarized via a Sentiment machine, tagging them as positive, negative or neutral.

The first main problem after collecting the data is how to analyze them, It is worthless to just look at the raw data, they don’t scale, they are too many.

What the software does is analyzing all the tweets with Machine Learning to get their sentiment and extract the most relevant keywords. These are some examples of the tweets given as input and the correlated output given by the program:

- "@HillaryClinton will receive the first question at tonight's presidential debate, according to @CBSNews #ClintonVsTrump".

Sentiment: Neutral.

- "Americans trust @realDonaldTrump to Make our Economy Great Again!".

Sentiment: Positive.

- "Racial discord was conceived, nurtured, refined & perpetuated by Americans incl @realDonaldTrump's father. Get real!".

Sentiment: Negative.

- "@wcve it's amazing how our city loves him, and he really loves our city. @HillaryClinton made a great choice for Vice President. @timkaine".

Sentiment: Positive.

Of course, these results are useless if we do not have a probability that the rating is actually correct, in other words, the accuracy of the machine. In the case of the analysis of Twitter, the short piece of text can be tricky to analyze also for a human analyst.

Starting from the big picture of this campaign, we can see from Tab. A, the overall sentiment towards the candidates. In that table are reported per each day the average sentiment, obtained by summing all the negative and positive tweets about one candidate.

As we can see, the overall sentiment of Trump is higher than Clinton's one, showing a first, unexpected in a certain way, sign of the final result. In general, it is also interesting to see how for both candidates there are way more negative tweets than a positive one.

In Tab B and Tab C we can see the number of tweets per each candidate, and also here we have a clear dominance by the Republican candidate.

Thanks to the software, is not only possible to compare the positive tweets or the negative one for each day, but also find the keywords used by the population in a specific timeframe. Let's analyze a day where the comments were particularly harsh: the 7th of October.

On that day, was published the videotape containing Trump's offensive comments about Alicia Machado, a Venezuelan-American actress

These are the most used keywords according to the software, which allows us to see the overall sentiment of that day, surely negative, but also why it is that negative. There are no doubts that Trump's analysts used these data to write his following speech and target directly the hottest sentiment expressed by the population.

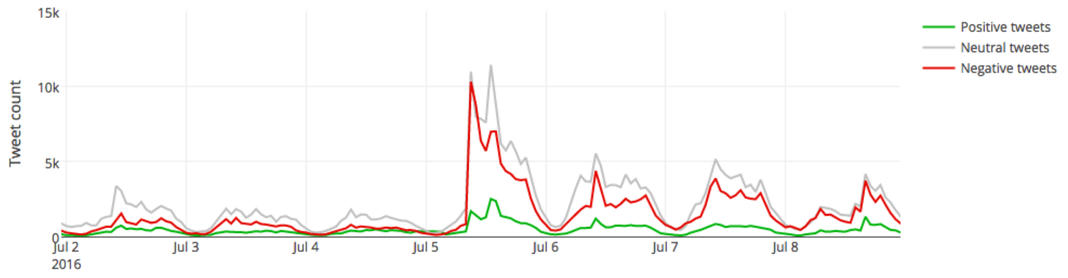
Raising our point of view on the specific case to have some general information, we got some useful insights which allow us to get a clear view of who was the favorite to win the election even before the 8th of November 2016

The first thing that stands out is that @realDonaldTrump gets mentioned much more than @HillaryClinton. Trump's Twitter presence is much larger than Clinton's. On an average day, Donald Trump's account gets about 450,000 mentions, while Hillary Clinton's account only gets 250,000. If it is true that "There is no such thing as bad publicity", as said by Phineas T. Barnum then Trump already had the first point here.

Secondly, out of those tweets, Trump has a better positive to negative ratio than Clinton. Even though most of the tweets were marked as neutral, for both candidates there are usually more negative than positive tweets. But a better positive to the negative ratio for Trumps leads to the conclusion that not only more people spoke about him, but also more people spoke positively about him, giving him a higher electorate base than Hillary, make him start from an advantageous position.

The power of the Sentiment Analysis machine is also that we can choose different levels of analysis that we can run, in particular, is interesting to watch also the reaction of American citizens on specific dates and specific keywords.

*July 5$^{th}$: the FBI says it's not going to end Clinton's email probe and will not recommend prosecution.*



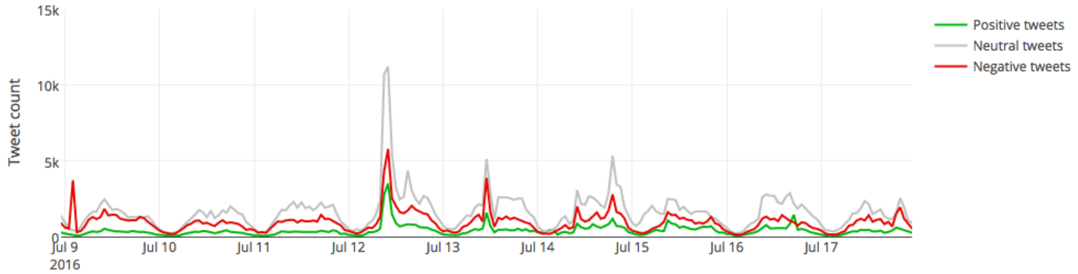

Tab D Clinton's tweet count by sentiment around July 5$^{th}$

On Tab D we can see the sentiment of candidate Clinton during the 5th of July. Additionally, to an unusual rise in traffic that day, a lot of people were not happy, and they were vocal about it. There's a significant rise in traffic on Clinton's side, with negative tweets taking off.

Checking out the keywords, you can see *emails* is a big one among the negative and neutral ones, alongside *FBI, dark day, American history*.

But, there's also a rise in positive tweets. A small one compared to the negative ones, but an increase nonetheless. Unfortunately, a lot of them seem to be sarcasm, so tweets like *thanks Obama* and *Great job with those emails!* are classified as positive.

*July 12$^{th}$: Bernie Sanders endorses Hillary Clinton for president*



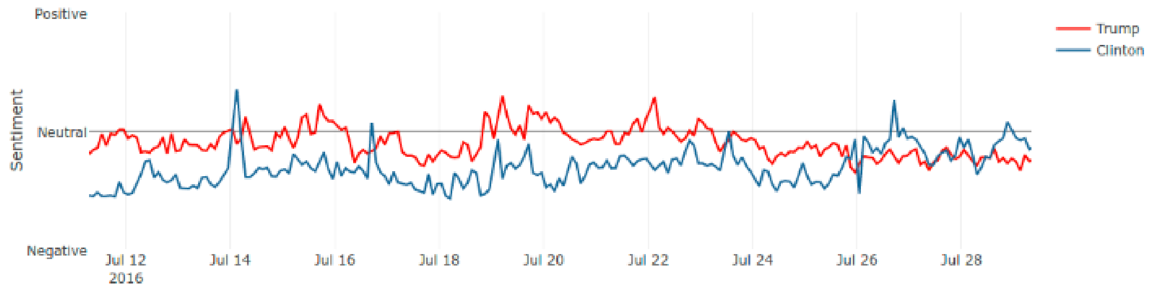

Tab E. Clinton's tweet count by sentiment around July 12$^{th}$

Looking at Tab E, there is a prominent peak in Clinton mentions that day, both neutral, negative and positive. However, the negative backlash was much smaller than the one on July 5th.

Clearly, this piece of news was met with mixed responses, with a rise in the positive tweets, keeping anyway a positive to negative ratio lower than 1, with Keywords like *Bernie supporters, thanks Bernie, best choice*, while negative Clinton keywords were bringing up criticism: *NAFTA, disastrous crime bill, email breach.*

It is worthless to say again what a useful piece of information was this for Clinton's crew, allowing to write the next speech, the next marketing campaign or political TV advertisement focused on fixing those sentiments.

*July 21st: Donald Trump accepts the Republican nomination:*

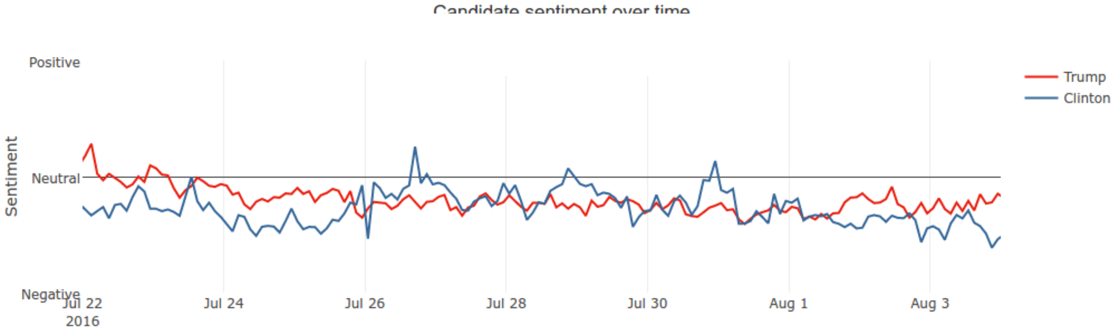



Tab F. Candidate Sentiment over time, expressed in positive tweets on negative tweets

During the Republican convention, which took place from July 18th to July 21st, there is a clear rise in Trump's sentiment, which means there were more positive tweets than negative ones, as expressed by Tab F.

An interesting thing that can be seen on this date is that *Ted Cruz* appears as a keyword on Trump tweets, but only in neutral and negative ones. Looking at the most relevant tweets, this means that people on Twitter were either reporting on the senator's speech or actively criticizing it, but not praising it. It's clear that people did not regard this as a good move by Trump's former rival.

*July 25$^{th}$ – 28$^{th}$: Clinton accepts the nomination in the DNC*




Tab G. Candidate Sentiment over time, expressed in positive tweets on negative tweets

This period is the only period where Clinton's sentiment is better than Trump's one. We can see that she had 2 positive peaks, the only two moments where the positive tweets were more than the negative ones.

Some of the relevant keywords for Clinton's positive tweets during these days includes *first woman president, proud, history first woman president, proud, history* and *amazing time tonight*, which clearly shows that people were excited and vocal about having the first woman nominee ever in the history of the US.

So far, it is clear to everyone that one of the two candidates is clearly advantaged in front of the other one in terms of citizens'sentiment.

Nevertheless, to have the complete pictures of what people felt during the elections and what was the real game changer of these race for the White House we must analyze the sentiment after some of the key debates.

*Sept 27th: Aftermath of the first presidential debate*

After the first debate, Clinton had the first glimpse of victory, which unfortunately will not continue until the end of the campaign, as history tells us.

After this debate, Clinton saw an advantage in terms of sentiment that lasted for a couple of days, returning around the 29$^{th}$ with the standard negative level lower than Trumps's level.

Positive tweets about Clinton praise her comebacks and mentions that she had won the debate. And on the other side, positive Trump tweets praise his job at the debate and also proclaims him as the winner.

*Oct 10th: Aftermath of the second presidential debate*

This time, there wasn't the same reaction as in the previous debate: even though Trump mentions were still about how Trump won, Clinton mentions mostly said that she lost. *Uncomfortable debate* is one of the top Clinton keywords, which says a lot about the public reaction to the second debate.

*Oct 19th: Aftermath of the third presidential debate*

Like in the first debate, the third and last presidential debate saw a considerable jump in the positive sentiment around Clinton, actually beating Trump in the general sentiment overall.

Some of the negative keywords for both candidates are related somehow to Clinton: on the Clinton, Twitter mentions we can find keywords like war drums, corruption, pathological liar and FBI documents. And on the Trump Twitter mentions we can find keywords like crooked Hillary, nasty woman and voter fraud.[42]

What we can conclude for this case study is that the sentiment is mainly decided by big events and it's the day to day that shapes public opinion in the end.

The conversation around the US elections has been omnipresent and highly polarized. Maybe it is not possible to predict for sure the results of the elections, but these tools help for sure the candidates and their teams to adjust their aim in order to get more supporters speech by speech.[20]

# 11. Customer Support Interaction Evaluation

As we mentioned before, customer support is nowadays vital to ensure a complete and quality customer experience, which starts from the moment of interest from a potential customer and ends with the customer support, that should last ideally until the end of the life cycle of our product, in other words when it is retired from the market.

Especially for a market like the mobile network one, customer support is crucial to emerge and beat the competitors. Everyone nowadays has a mobile phone, following that everyone has a mobile network provider and a contract with them.

Luckily for the consumers, less for companies, the cost to switch from a company to another has decreased constantly and reached nowadays a cost close to zero. At the same time, the process became easier, leading the companies to focus a lot of efforts and resources to Customer Retainment.

In this case study, we will use the Sentiment Analysis to understand how customers react to various customer services of the four biggest companies of the network: AT&T, Verizon, Sprint and T-Mobile.

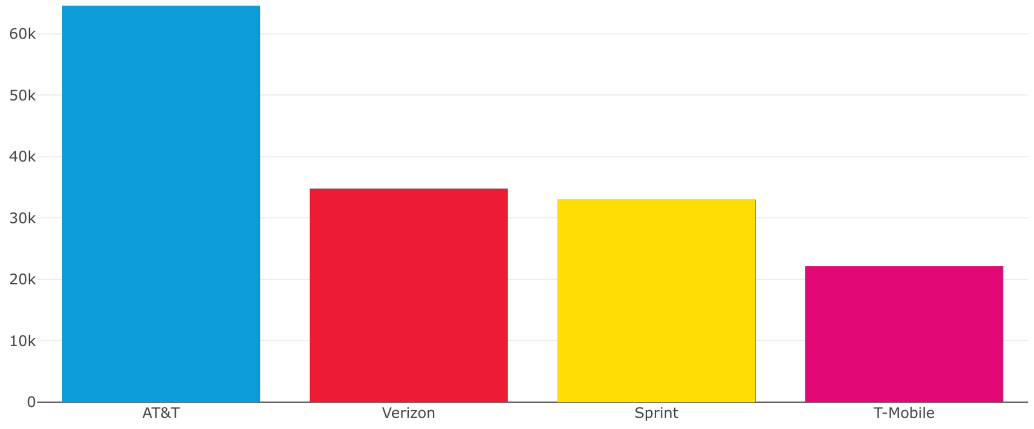
The focus is on the overall reactions to the customer support, why they reacted like that and what a company can do to improve the customer experience and, in the end, retain or attract more customers.

We are seeing new trends in customer support. Some companies are starting to have a different social media appearance, trying to appear more friendly and close to the customer lifestyle, instead of using the professional-servile approach.

Some signs of this approach are every member of the social media team signing with their name, and engaging in conversations with users that don’t involve the company at all. These conversations can even involve friendly banter, which was absolutely unthinkable in the old days.

The Sentiment Analysis was carried by collecting tweets mentioning one of those four big carriers with positive, neutral or negative labels.

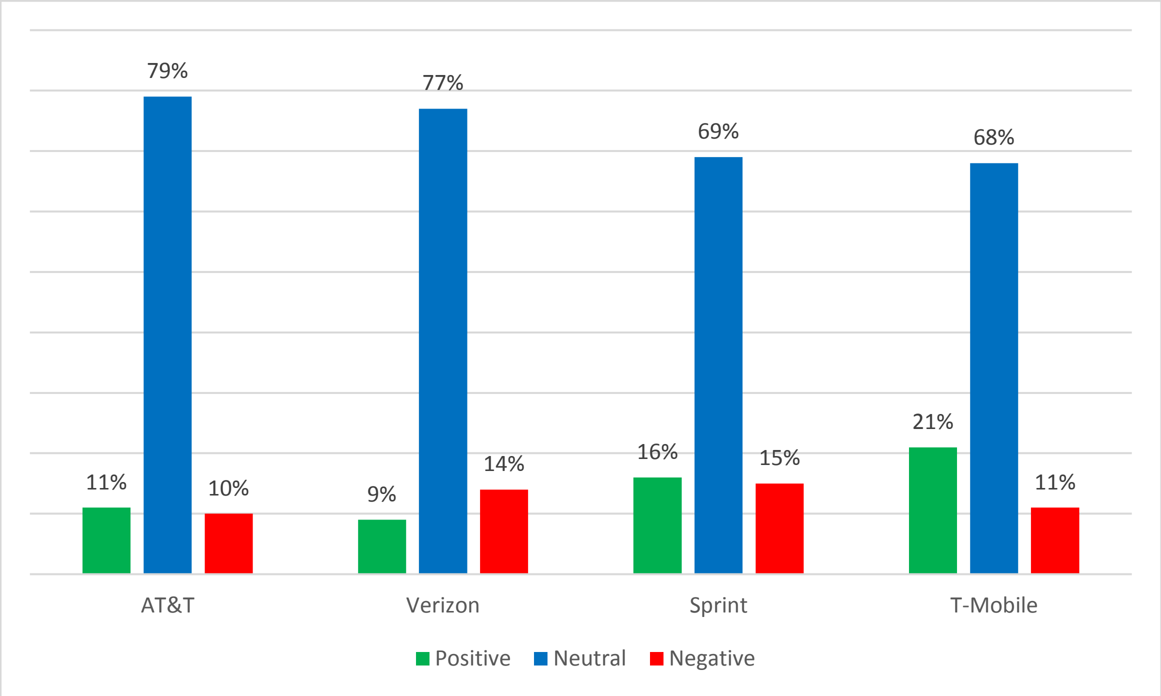
Starting with some data, on a given week the average number of tweets mentioning a company (either by name or by the handle) has AT&T on top, with an average of 64,000 tweets a week.




Tab A: Average mentions per week per company

This was pretty surprising on its own, especially considering that Verizon and AT&T have almost the same number of subscribers. Also, Verizon has a lot more followers than AT&T, yet they have half the number of tweets on any week. This by itself is interesting, since one would expect carriers with a similar number of subscribers (or Twitter followers) to have the same number of mentions.

Moving from raw numbers to the sentiment linked to those numbers is interesting to see the composition of the labels for those tweets.




Tab B: Composition of tweets per label per company

On the positive side, clearly on top comes out T-Mobile, with 21% of the tweets mentioning them being positive. They are ahead by their next competitor, Sprint, by 5 percentage points and have double the rating than the larger companies, AT&T and Verizon.

Interestingly, the larger companies have less positive tweets than the smaller ones, and this fact is probably due to the more focused and tailored customer care that a company can deliver with a smaller customer base.

Now on the neutral side, turns out most tweets are actually neutral tweets: these are mostly factual questions, answers, or opinions that don't express a sentiment.

Here, the largest carriers have the bigger chunk of their tweets with neutral sentiment. So far what we are seeing is that a larger percentage of people are simply talking about them, without expressing a sentiment.

What is really surprising is that 3 out of 4 of these companies have a positive to negative ratio above 1, meaning that they receive more positive tweets rather than negative ones.

Considering that we are talking about social media platforms, in particular, Twitter, contain mostly negative tweets and fights.

The only exception is Verizon: they are the only ones that have more negative tweets than positive ones. This would mean that as far as Twitter sentiment goes, Verizon is seen as the worst carrier of all.

As for the others, AT&T has about the same for both tags, while T-Mobile is the one that fairs the best. Sprint is ahead of the pack in negative tweets 15%, which combined with their 14% of positive tweets indicates that it's the company where people have a strong opinion, either positive or negative.

Even though this data is interesting for consumers, allowing them to select the carriers with the best customer support or the best relationship with their clients, can be used by the companies just to position themselves towards the principal competitors.

The Sentiment Analysis tool, allow also to understand what these tweets contain, what do the positive tweets have in common, what the negative ones, why one provider has more positive tweets than the others and what they can do to not lose market share.

Starting with the negative tweets, it is curious to notice that all the companies have the same common complaints: *bad customer service, bad reception, high prices.* People especially complain on Twitter about bad customer service, probably because it is easier to be heard on this platform.

Now, there were some insights that only appeared in one of the four carriers, complaints unique to a single carrier. For example, T-Mobile was the only one where people complained about the quality of their LTE service (the US transposition of our 4G network), while Verizon was the only one where people complained about their 'Unlimited plan" which, according to what emerged from the analysis, apparently isn't unlimited.

Moving to positive tweets, it is interesting to compare the two carries at the top and at the bottom of the positive tweets ranking: T-Mobile and Verizon.

Verizon’s keywords were something to be expected: things like *better network, new phone, rewards, thanks, quality customer service* and so on. The related sentences were regular interactions between the guys at support and the customers.

Now, looking at T-Mobile, the picture changes. Many keywords here are the names of the members of the customer support team — indicating that users are engaging highly with them and addressing them on a first name basis.

Looking at their tweets with replies page, it’s clear that their game is on another level, with gif replies and emojis being common, and engaging on seemingly unrelated topics with users, going from football games to the latest episode of a TV series.

Compare with Verizon’s tweets: very dry and professional, the expressions of sentiment feel forced and almost fake. This really makes a difference with their customers.

# Conclusion

The main aim of this document was to understand the state of art of the Sentiment Analysis and the potential behind this new field of the Machine Learning world.

Application for this technology are countless, just limited by the skills and the creativity of each individual. I brought the application in advertisement, marketing, politics, just because the flexibility of Sentiment Analysis is infinite.

The only limit we must acknowledge about Sentiment Analysis is that, as any new technology does, it will for sure encounter resistance at the implementation and, even after the implementation, requires time to get understood and to leverage the return on investment.

Therefore, and don't see it applicable, at least in the short term, inside Vicenza or even in the region, due to the micro dimension of the firms in our economic environment.

Specific skills and initial investments are required to make it work and, even though Sentiment Analysis can be carried out even with small amount of data, it gives the maximum utility with data that cannot be handled by human beings.

This amount of data is not accessible to small enterprises, that is why I think the main application will be carried out from Medium and Big Companies, with the knowledge and the resources.

# BIBLIOGRAPHY

[1] Anatomy of Facebook, Lars Backstrom

[2]Statista,https://www.statista.com/statistics/562397/worldwide-revenue-from-social-media/

[3] Danah M. Boyd, Nicole B. Ellison, Social Network Sites: Definition, History,

and Scholarship, 2008

[4] Haythornthwaite, (2005). Social networks and Internet connectivity effects. Information, Communication, & Society

[5] Sunde´n, J. (2003). Material Virtualities.

[6] Festa, P. (2003, November 11). Investors snub Friendster in the patent grab.

[7] Chafkin, M. (2007, June). How to kill a great idea!

[8] Shirky, C. (2003, May 13). People on the page: YASNS.Corante's Many-to-Many

[9] Boyd, d. (2006, March 21). Friendster lost steam. Is MySpace just a fad? Apophenia Blog.

[10] Perkel, D. (in press). Copy and paste literacy? Literacy practices in the production of

a MySpace profile

[11] BBC. (2005, July 19). News Corp in $580m Internet buy.

[12] Madhavan, N. (2007, July 6). India gets more Net Cool. Hindustan Times.

[13] https://www.brandwatch.com/blog/amazing-social-media-statistics-and-facts

[14] (Boyd, Dana; Crawford, Kate (21 September 2011). "Six Provocations for Big Data". *Social Science Research Network: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*.)

[15] *"Data, data everywhere". The Economist. 25 February 2010.*

[16] “All too much” The Economist, 25 February 2010

[17] comScore/the Kelsey group. Online consumer-generated reviews have a significant impact on offline purchase behavior. Press Release, November 2007. http://www.comscore.com/press/ release.asp?press=1928.

[18] John A. Horrigan. Online shopping. Pew Internet & American Life Project Report, 2008.

[19] Spiegel Research Center, How Online Reviews Influence Sales/Evidence of the Power of Online Reviews to Shape Customer Behavior, 2017

[20] Bing Liu. *Sentiment Analysis and Opinion Mining*, Morgan &Claypool Publishers, May 2012

[21] https://monkeylearn.com/sentiment-analysis/

[22]https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/

[23] http://sentiment.christopherpotts.net/stemming.html

[24] Dr. S.Vijayarani, R.Janani Text Mining: Open Source Tokenization Tools – An Analysis

[25] Lemmatization and POS Tagging for Deep Learning, “Bitext, 28/02/2018”

[26] Alvin Chung, Supervised Machine Learning—Linear Regression in Python, 2017, Hackernoon.com

[27] Prof. L. Argerich, “What is overfitting?”, Quora 30 June 2016

[28] L. Zhang, Deep Learning for Sentiment Analysis: A Survey

[29] Langacker, Corpus Linguistic Methods – A practical introduction with R and Python

[30] “The world’s most valuable resource is no longer oil, but data” The Economist, 06/05/2017

[31] https://dictionary.cambridge.org/dictionary/english/irony

[32] https://dictionary.cambridge.org/dictionary/english/sarcasm

[33] Gaël Guibon, Magalie Ochs, Patrice Bellot, 2016, From Emojis to Sentiment Analysis

[34] C. Kelly. Do you know what I mean >:(: A linguistic study of the understanding of emoticons and emojis in text messages.

[35] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welpe, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment"

[36] Forbes, Mar 2012, "If You're Not Paying for It, You Become the Product"

[37] Hern, Alex (April 10, 2018). "How to check whether Facebook shared your data with Cambridge Analytica". The Guardian. Retrieved July 3, 2018.

[38] Y. Liu et al., 2007, "ARSA: a sentiment-aware model for predicting sales performance using blogs"

[39] M. McGlohon, N. Glance, Z. Reiter, 2010, "Star Quality: Aggregating Reviews to Rank Products and Merchants"

[40] McKinsey, December 2017, Smart Cities: Turning Opportunity Into reality

[41] McKinsey Quarterly, August 2016, "The CEO guide to customer experience

[42] Bruno Steccanella, MonkeyLearn Blog, Oct 2016, "Trump vs Hillary: sentiment analysis on Twitter mentions"

[43] Bruno Steccanella, MonkeyLearn Blog, Oct 2017, "Analyzing customer support interactions on Twitter with Machine Learning"

[44] Statista, Social network penetration worldwide from 2014 to 2021

[45] S. Porio et al., 2013, "Dependency-Based Semantic Parsing for Concept-Level Text Analysis"

[46] The Globe and Mail, January 2014, "No strings, please: Expedia Canada and falls flat"

[47] Dayne Shuda, Ghost Blog Writers, Nov. 2012, “How Long Does a Tweet Last?

[48] Proof Strategies, “Advertising on Twitter: Promoted tweets vs Promoted trends”

[49] H Saif, M Fernandez, 2013 “Evaluation datasets for Twitter Sentiment Analysis: a survey and a new dataset, the STS-Glod”

[50] Netpromoter.com “What is Net Promoter”