Instituto Universitário de Lisboa

Science and Information Technology Department

Sentiment Analysis in Retail: The Case of Parfois Facebook Page

António José dos Santos Teixeira

Dissertation submitted as partial fulfilment of the requirement for the degree of Master in Informatics Engineer

Supervisors

Prof. Raul Manuel da Silva Laureano, Assistant Professor, ISCTE Business School, Department of Quantitative Methods for Management and Economics

Prof. Luis Miguel da Silva Laureano, Assistant Professor, ISCTE Business School, Department of Finance

October 2017

“Knowledge is of two kinds. We know a subject ourselves, or we know where we can find information on it.”

— Samuel Johnson

Acknowledgments

As all things in life this thesis come to life with the help of a number of people. At the moment this research is finished, and looking back on the process, I feel both the need and the responsibility to thank everyone who, directly or indirectly, helped me to achieve the final results that are presented in this research.

First of all, I would like to express my deepest and sincerest gratitude to Professor and Supervisor Raul Laureano for his help, advice, guidance and teachings in so many subjects. His help was paramount in writing this thesis.

I would like to thank to my wife and my son. Although they endured my bad humor and lack of attention during the writing of the thesis, they always returned the smiles and support on which I leaned to complete it.

To my mother and sister, for all the support along the past years my special thanks. The road was that much easier with your help.

A special thanks to Engineer Luis Viera for his help revising the English of this research, to Engineer Ricardo Luis Moura for his help answering some questions about the theme of this research and Engineer José Ribas as he was the person who first gave me the theme for this research.

Last, but by no means least, I cannot fail to mention the star that lit up in the sky on July 22, 2002. Thanks to him, I am the person that I am.

Abstract

The way that consumers are interacting with brands is changing, and in Retail it is no different. With the growth of internet usage and with all the social networks that we interact with, social media is gaining more and more relevance and importance. This research extracted 1.845 posts, 8.256 comments and more than 500.000 reactions from Parfois Facebook page. The comments were translated to English due to having comments made in several different languages, modelled and finally made the sentiment analysis. This analysis was made concerning the post dates, the reasons of the post and the products associated in the post.

It was used decision tree algorithms to predict sentiments, so it can be predicted the sentiment when making a new post.

With the Sentiment Analysis from Social Media, Parfois can gain understanding about their own brand, from the marketing department through to the buying or even design departments. Using Social Media analysis together with Business Intelligence, can help Parfois decision makers gain competitive advantage regarding their competitors or even improve their products.

Keywords: Sentiment Analysis, Social Media, Facebook, Retail, Marketing

Resumo

A maneira como os consumidores interagem com as marcas está a mudar, e no retalho não é diferente. Com o aumento do uso da internet e com todas as redes sociais que interagimos, as redes sociais ganham mais relevância e importância. Esta pesquisa extraiu 1.845 posts, 8.256 comentários e mais de 500.000 reações da página de Facebook da Parfois. Os comentários foram traduzidos para o inglês devido ao fato de haver comentários feitos em várias línguas diferentes, modelados e finalmente feita a análise de sentimentos. Esta análise foi feita em relação às datas das publicações, os motivos do post, os produtos associados ao post.

Foram utilizados algoritmos de árvores de decisão para prever sentimentos para que se possa prever o sentimento ao fazer um novo post.

Com a analise de sentimentos das redes sociais, a Parfois pode entender melhor a sua própria marca, desde o departamento de marketing até ao departamento de compras ou mesmo o departamento de design. Usar a análise de sentimentos das redes sociais junto com o Business Intelligence organizacional, pode ajudar os decisores da Parfois a ganhar vantagem competitiva em relação aos concorrentes ou mesmo a melhorar seus produtos.

Palavras chave: Análise de Sentimentos, Social Media, Facebook, Retalho, Marketing

Abbreviations

ANN – Artificial Neural Network

AUC – Area Under the ROC Curve

BI – Business Intelligence

BM – Brand Management

CART – Classification and Regression Tree

CHAID – Chi-squared Automatic Interaction Detector

CRISP-DM – Cross-Industry Standard Process for Data Mining

DM – Data Mining

EWOM – Electronic Word-Of-Mouth

KPI – Key Performance Indicator

NER – Stanford Named Entity Recognizer

ROC – Receiver Operating Characteristic

ROI – Return on Investment

SA – Sentiment Analysis

SM – Social Media

SMA – Social Media Analytics

SMD – Social Media Data

SMN – Social Media Networks

SMP – Social Media Platform

SNS – Social Network Site

TM – Text Mining

Table of contents

1 INTRODUCTION ... 1

1.1 Theme and Motivation ... 1

1.2 Goals and Contributions ... 2

1.3 Methodology Overview ... 3

1.4 Structure ... 4

2 LITERATURE REVIEW ... 5

2.1 Social Media ... 6

2.2 Social Media in Retail ... 15

2.3 Sentiment Analysis ... 19

The importance of Facebook? ... 22

2.4 Sentiment Analysis Studies in Retail ... 24

2.4.1 Cosco and Walmart ... 24

2.4.2 Facebook posts ... 26

2.4.3 Burberry case ... 27

2.4.4 Louis Vuitton case ... 29

2.5 Business Intelligence ... 30

2.5.1 Social Business Intelligence ... 30

2.5.2 KPI ... 31

2.6 Social media metrics ... 33

2.7 Brand Management ... 38

Brand Analysis ... 39

3 METHODOLOGY ... 41

3.1 Business Understanding ... 42

3.2 Data Understanding ... 44

3.2.1 Data extraction procedure ... 45

3.2.2 Attributes collected ... 49

3.3 Data Preparation ... 52

3.3.1 Emoticons and emojis ... 53

3.3.2 Comments translation ... 54

3.3.3 Text analysis to retrieve sentiments and types ... 55

3.3.4 Other transformations ... 58

3.3.5 Attributes for modeling ... 60

3.4 Modeling ... 63

3.5 Evaluation ... 66

3.6 Deployment ... 68

4 RESULTS AND DISCUSSION ... 69

4.1 Comments characterization ... 69

4.1.1 Contents ... 70

4.1.1 Sentiments *71*
4.1.2 Relationship between content and sentiment *73*
4.2 IDENTIFICATION OF THE CHARACTERISTICS OF THE POSTS THAT INFLUENCE THE ASSOCIATED SENTIMENT OF THE POST 76
4.2.1 Posts characterization *76*
4.2.2 Evaluation of both the sentiment and the interaction level associated to the Posts *80*
4.2.3 Relationship between products associated to the posts with the interaction level *81*
4.2.4 Relationship of the products associated to the posts with the sentiment of the post *84*
4.2.5 predictive model of the positive sentiment associated to the post *87*
4.3 DISCUSSION 92
5 CONCLUSION 97
5.1 SUMMARY 97
5.2 CONTRIBUTIONS 97
5.3 LIMITATIONS 99
5.4 FUTURE WORK 99
6 REFERENCES 101

List of Tables

Table 1: Fortune 500 adoption density of social media platforms by industry 6
Table 2: How virtual customer environments create value 11
Table 3: Tesco, Amazon and Wal-Mart: social media activity 16
Table 4: Average number of fans / followers in retailer's SMP, in Portugal (2016) 17
Table 5: Average number of admin posts made by retailers in SMP, in Portugal (2016) 17
Table 6: Average number of interactions in retailer's SMP, in Portugal (2016) 18
Table 7: Average response rate in retailer's SMP, in Portugal (2016) 18
Table 8: Interaction types on Facebook in retailer's page, in Portugal (2016) 18
Table 9: Facebook post types made by retailers in their Facebook page, in Portugal (2016)..... 19
Table 10: Interaction types on retailers Twitter, in Portugal (2016) 19
Table 11: Facebook terminology 24
Table 12: Cosco Twitter data from december 1, 2014 – February 29, 2015 25
Table 13: Walmart Twitter data from december 1, 2014 – February 29, 2015 25
Table 14: Post reason distribution 26
Table 15: Post categories 27
Table 16: Kpi's examples 33
Table 17: Metrics used by brands 34
Table 18: Evaluation range examples 35
Table 19: Facebook insights on page level 36
Table 20: Facebook insights on post level 37
Table 21: Post attributes collected 51
Table 22: Reactions attributes collected 51
Table 23: Comments attributes collected 52
Table 24: Part of the dictionary of emoticons and emojis 53
Table 25: Dictionary of types of comments' content 56
Table 26: Sentiment categories used in IBM SPSS Modeler 57
Table 27: Categories and sub-categories used 58
Table 28: Variable transformation 59
Table 29: List of products attributes for modeling 60
Table 30: List of post attributes for modeling 61
Table 31: List of date post attributes for modeling 62
Table 32: List of fans / follower's reactions for modeling 62
Table 33: List of reactions for modeling 63

Table 34: CART models used ... 64
Table 35: C5 models used ... 65
Table 36: Models parametrizations ... 65
Table 37: Classification matrix ... 66
Table 38: Types of contents distribution ... 71
Table 39: Comments sentiment distribution ... 72
Table 40: Comments sentiments distribution by year ... 72
Table 41: Sentiment distribution by type of content ... 75
Table 42: Sentiment distribution by comments' language ... 76
Table 43: Number of posts by date attributes ... 77
Table 44: Post characterization ... 78
Table 45: Distribution of the products associated to posts ... 79
Table 46: Evaluation of sentiments and public involvement ... 80
Table 47: Distribution of the interaction attributes ... 81
Table 48: Interaction metrics by products associated ... 82
Table 49: Post sentiment present by products associated ... 85
Table 50: Post sentiment by products associated ... 86
Table 51: Evaluation metrics for CART algorithm ... 87
Table 52: Evaluation metrics for C5 algorithm ... 87
Table 53: Evaluation metrics for ANN algorithm ... 88

List of Figures

Figure 1: Different types of social media for Kaplan & Haenlein 8
Figure 2: Different types of social media 9
Figure 3: Sentiment Analysis topic interest over time 20
Figure 4: Most popular social networks.. 22
Figure 5: Social media platform with the higher number of active users worldwide 23
Figure 6: CRISP-DM Methodology.. 41
Figure 7: List of countries where Parfois have stores 43
Figure 8: Relational database 48
Figure 9: Parfois Post from Facebook 50
Figure 10: Comments made in Parfois post 50
Figure 11: ROC Curve representation 67
Figure 12: Comments sentiments distribution by year 69
Figure 13: Comments distribution by language used 70
Figure 14: Global Word Cloud 70
Figure 15: Global network linking contents and sentiments... 73
Figure 16: Network linking contents and negative sentiments 74
Figure 17: Network linking contents and positive sentiments 74
Figure 18: ROC of the three predictive models 88
Figure 19: Rule set for the best predictive model of the positive sentiment (C5 model)........... 89
Figure 20: Predictors importance (target = positive sentiment)... 89
Figure 21: Distribution of the predicted sentiment by post reason 90
Figure 22: Distribution of the predicted sentiment by month of the year 91
Figure 23: Distribution of the number of products associated by predicted sentiment 91
Figure 24: Suggestion of dimension and fact tables 100

1 INTRODUCTION

In this first chapter is presented the theme, its relevance and delimitation. After are described the objectives and the methodology. Finally, all the structure and practical work of the thesis is explained.

1.1 THEME AND MOTIVATION

With the growth of internet availability, it grows the amount of data produced every day. Every second we have 6.000 tweets on Twitter (Internet live stats, 2017). Whether we like it or not, nowadays the information is spreading in the speed of a post or a tweet and this data is consumed by us at every second, and can drive us to make decisions, good or bad. Brands are no different and they are improving and putting more efforts in this area, in order to make good decisions.

The consumers gain more and more influence with the information that comes from Social Media (SM) placed in the Social Network Sites (SNS) used by brands. When using the SNS and posting information about their products or services, the brands allow fans / followers to comment about them, so they can express their sentiments. These sentiments can be analyzed and applied in the retail and help in the brand management. With all this information accessible by everybody, negative sentiments can be very harmful to the brand reputation.

Even when we ask ourselves how many users are following us in Tweeter, or how many likes we have in a certain post in Facebook, we are analyzing sentiments and using it in our own Business Intelligence (BI). Brands are also starting to do it, to check what people think about the brands, how to innovate, and what to change. Zeng *et al*. (2010:14) concluded that "*for-profit businesses are tapping into social media as both a rich source of information and a business execution platform for product design and innovation, consumer and stakeholder relations management, and marketing.”*.

In the present, consumers tend to join network communities that have same tastes as them. These communities are called SNS and are defined by Boyd & Ellison (2007:210) as “a 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.”.

With the increase of Social Media Networks (SMN) and the increase of the usage of the SMN, it exists a high volume of data created by fans / followers. All the data produced is available to be extracted and analyzed by anyone, as it is public. This means that the information can be analyzed by the brand or by the competitors. Analyzing the comments and the reactions of a SMN, the competitors can realize what are the strength and the weakness of their competitors and use that information to gain competitive advantage or even to pass a negative sentiment about competition to their fans / followers or future costumers.

Therefore, brands that have online platforms, must have the concern to know what their clients or future clients speak and think about them, even before getting to know their competitors or other external stakeholders. It is with this perspective that this research will try to answer to the question: “is it possible to extract useful information to the business from Facebook?”.

1.2 GOALS AND CONTRIBUTIONS

To answer the questions, two main goals are established, each of them with specific objectives:

O1: Characterize the comments made in brand Facebook page by the fans / followers of the brand Facebook page.

O1.1: Describe the content of the comments;

O1.2: Evaluate the sentiment associated to the comments;

O1.3: Analyze the relationship between the content and the sentiment;

O2: Identify the characteristics of the brand posts that influence the associated sentiment and interaction level by fans / followers of the brand in brands Facebook page;

O2.1: Characterize the posts;

O2.2: Evaluate the sentiment and interaction level associated to the posts;

O2.3: Analyze the relationship between products associated to the posts and the interaction level;

O2.4: Analyze the relationship between products associated to the posts and the sentiment of the post;

O2.5: Create a predictive model of the positive sentiment associated to the post.

The achievement of these objectives brings contributions to brand professionals by allowing them to answer questions such as:

- what products leads to a positive sentiment towards to brand?;
- what products have more fans / follower's interactions and what type of reaction;
- are the metrics adequate to Parfois?.

Moreover, we can help the professionals to get to know the sentiments of their costumers, current and future ones. We can help them recognize the efficiency of the use of SMN, in particular, the use of Facebook in retail. Another contribution for professionals will be what type of posts generate more likes and mores positive sentiments.

Additionally, this research contributes to the literature when presenting a case of an international case in which, in addition to evaluating the feelings and reactions of fans / followers, we identify the explanatory factors of these feelings, namely at the level of the characteristics of posts, allowing to manage trends in Facebook.

1.3 METHODOLOGY OVERVIEW

In this research, we used CRISP-DM (Cross-industry Standard Process for Data Mining) methodology. This methodology is composed of six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Chapman *et al*., 2000).

For this research, we used the data collected in Parfois Facebook page. Parfois is a Portuguese brand, who creates and sells women accessories. Although the brand uses several SMN, Facebook is the one where Parfois is more active and have more interactions with the fans / followers. All the posts made in Facebook were collected since the first post was made in the brand page. Therefore, we analyzed 1.845 posts (from 2010 to 2017), 8.256 comments and 519.336 reactions.

To extract the sentiments and content types from the comments we used text analytics tools, transforming non-structured data into structured data. Then, this data is analyzed using descriptive statistics techniques, and to create a predictive model, decision trees and artificial neural networks were used. Moreover, whenever possible, we used open source tools to extract and transform data, while for the statistical analyses IBM SPSS Statistics and IBM SPSS Modeler were used.

1.4 STRUCTURE

This dissertation has a traditional structure. Besides this introduction the dissertation contains a Literature Review in chapter 2 with an overview about sentiment analysis and how retail can use it. In chapter 3 we have the Methodology used to extract and analyze the sentiments in a retail Facebook Page. In chapter 3 it is also described the CRISP-DM methodology and all the work made in the several phases of the methodology. In chapter 4, Results and Discussion, it is presented and discussed the findings of the research. In the last chapter, Conclusions, it is described the conclusions and it focus the contributions of the research.

2 LITERATURE REVIEW

Like in all industry sectors also retail needs to combine a lot of factors to be sustainable. The competitors are fierce and nowadays all companies need to be totally focused even in some details, like social media. Almost all brands are starting to have Facebook, Instagram, Twitter or Pinterest (Pletikosa and Michahelles, 2011) and the way they collect and analyze the information can give them advantage against their competitors (Dey *et al*., 2011).

Brands give high importance to the opinion and sentiment of the fans / followers (D'Andrea *et al*., 2015). The data collected in Social Media Platform (SMP) can be a guide to make some important decisions and even can compare the customers' sentiments among the competitors, about the trends, products or even services (Wu *et al*., 2015).

Beside the decisions that can be made with the information collected and analyzed in the SMP, another challenge is to create trends, or even change the trend to another direction. The diverse tastes of fans / followers present a challenge too, because it is getting harder to know or drive all trends that are appearing (Asur and Huberman, 2010).

All of these decisions that can be made due the information collected, analyzed, prepared and presented to decision makers can give a competitive advantage against the competitors. And it is how the concept of Competitive Intelligence appears.

Competitive Intelligence is defined as "a combination of defining, gathering and analyzing intelligence about products, fans / followers, competitors and any aspect of the environment needed to support executives and managers in making strategic decisions for an organization. As opposed to industrial espionage, competitive intelligence is viewed as a legal business practice with focus on the external business environment." (Dey *et al*., 2011: p. 1). This competitive intelligence can give information about the company and even can give information about competitors, and can be used to prepare reactions to threat or spot weaknesses and transform them into strengths (Ruhi, 2012; Power *et al*., 2011).

The huge amount of raw data that crowds the internet and SMP can be gathered, collected and analyzed using algorithms. The data collected can be used for different purposes and for different types of industries (Culnan *et al*., 2010) as we can see in Table 1. We can verify that besides the IT sector that has the higher average value of Social Media Applications use, the sector which has the second higher value is the retail sector.

Table 1: Fortune 500 adoption density of social media platforms by industry

Industry (No. of Firms)	**Number of Social Media Applications per Firm**					
	0	***1***	***2***	***3***	***4***	***Average***
Distribution (23)	11	10	1	0	1	0.70
Energy (78)	43	25	9	1	0	0.59
Financial Services (72)	26	13	22	8	3	1.29
General Services (38)	17	7	6	8	0	1.13
Healthcare (18)	5	8	4	0	1	1.11
IT (40)	2	7	4	10	17	2.82
Manufacturing (146)	61	31	30	18	6	1.16
Retailing (54)	10	8	18	13	5	1.91
Transportation (17)	1	5	8	2	1	1.82
Other/Misc. (14)	5	2	6	1	0	1.21
Total (500)	181 (36%)	116 (23%)	108 (22%)	61 (12%)	34 (7%)	1.3

Source: Culnan et al. (2010 : p. 245)

In this chapter, we present SM and its uses within the retail industry. Then, an overview of a specific analysis in the scope of SM – sentiment analysis (SA) is presented. After, we give an overview about studies of SA in retail and its application in the retail sector using BI for the brand management (BM) in order to gain competitive leverage.

2.1 SOCIAL MEDIA

SM is not a new thing. It started a long time ago with written correspondence, taking other forms like telegraph, phone or even the word of mouth as technology evolved (Hendricks, 2013).

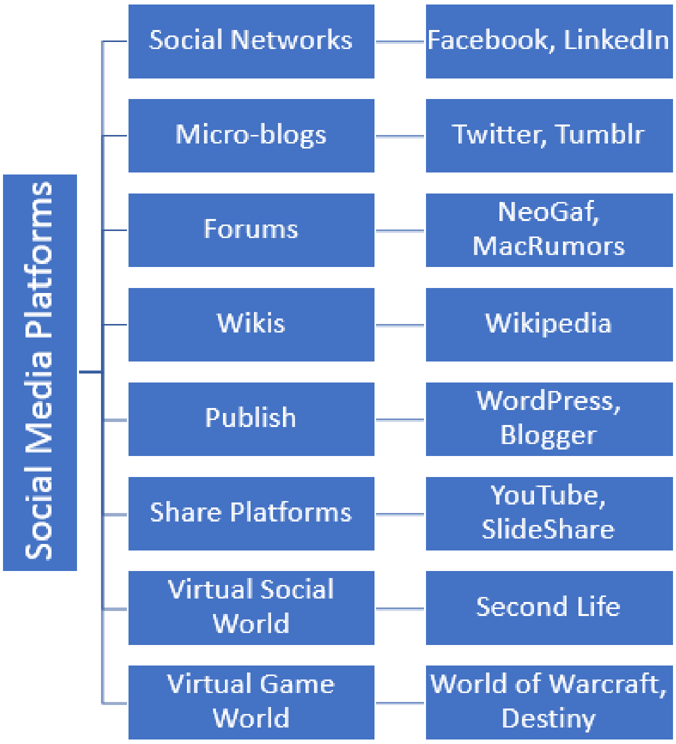
The advent of the internet and its expansion, brought a boom of many SMP like Facebook, Twitter, Instagram, LinkedIn, Google+, as well as others not so well known (Kallas, 2017). SM in the present context can be defined as "*... web-based platforms that support online social networking, community-building and maintenance, collaborative information production and* *sharing, and user-generated content production, diffusion and consumption.*" Fuchs (2011: p. 3).

Additionally, Social Media Analytics (SMA) is defined as “advanced informatics tools and analytics techniques to collect, monitor, and analyze social media data to extract useful patterns and intelligence.” (Zeng et al., 2010: p. 1623). There is no consensus definition between so many authors. For Fan and Gordon (2014) SMA is as interdisciplinary modeling and analytical paradigm that consists of three steps: 1) capture data from various courses; 2) understand data using various analytics and models; and 3) summarize and present the findings for decision making. SMA involves analysis, management and visualization of the similar types of datasets from online activities of consumers (Kiron, Perguson, & Prentice, 2013). And Liu (2012) identify SMA as the various analyses of people’s opinions, sentiment, evaluation, attitude, judgments and emotions towards various objects, including issues, products, services, organizations or individuals. Finally, development and the evaluation of scientific methods, technical frameworks, and software tools to track, model, analyze and mine large-scale Social Media Data (SMD) is the definition of SMA of Stieglitz et al. (2014).

SMA is the whole collection of information we have access in all online tools, or in other words, all the interactions users make online like sharing, extracting or either exchange experiences or sentiments with a community that have the same purposes or affinities (Zeng *et al.*, 2011). Even more, consumers are using the SMP of companies to get information about products, to complain, or just making some comments (Hennig-Thurau *et al.*, 2010). The social networks interactions are changing between companies and consumers.

What we can realize when we make a search in the internet is that there is no consensus among authors and/or bloggers regarding the name or categories of the social media types. Kaplan & Haenlein (2010) make a different classification of the social media types. The authors classify the social media types in eight types (Figure 1).

Figure 1: Different types of social media for Kaplan & Haenlein



Source: adapted from Kaplan and Haenlein (2010)

What is important to retain about SM, is that it is changing and evolving with time, adapting to trends, needs and wishes of its users. SMA is a point where all type of companies tends to converge. For example, Facebook started off to be more social, where we could write ideas or thoughts. Now we can already share media, and even can buy products from companies. So, with the evolution of technology, SM is changing and adapting too. What was only exclusive for each type of SM, are now shared functionalities mixed between them. Figure 2 shows the different types of social media (Baldwin, 2015).

Figure 2: Different types of social media

Source: Baldwin (2015)

So, the most usual social media types are:

- Social networks: These have the purpose of socialization with friends, classmates, co-workers or just people with shared ideas or goals. Inside the social networks, we can have networks with different objectives like, for example, social, professional, informational, educational, hobbies or even academic. Usually in these kind of networks, we find the target person, send an invitation to add her to the network and then interact by reading her posts, checking what she is doing with real time updates. Examples of this category are Facebook, Twitter or LinkedIn (Musiał & Kazienko, 2013);
- Discussion forums: These are the way to connect people with the same interests. A forum is a website where people interact with each other engaging in discussions, opinion giving or even advice trading in order to help each other. Usually, each forum has its own subject (cars, books, movies, sports, video-games, etc.). In this category we have Reddit, Quora, Digg (Curtis, 2017);

- Bookmarking: In these sites, we provide links that we "digg" or search in order to assemble them in one subject. This set of data is then shared with others to help them to "digg" more or even to help others that have the same interest in that subject. Pinterest and Flipboard belong to this category (Thompson, 2017);
- Media sharing networks: Not all networks are for reading. We have networks like the well-known YouTube where we can watch videos of all sort like tutorials, educational, films, comedy, television episodes and sports. These types of networks allow comments, rating, friends, and channels. In this type of SMP, we have Instagram, SnapChat or Youtube (Thompson, 2017);
- E-commerce: The new trend now is the ability to search, view and buy products in the social media. We can aggregate in the sites different products from distinct retailers and sell them, in an online marketplace. Here we have Farfetch, Polyvore, Etsy, Fancy (Thompson, 2017);
- Personal networks: Platforms where people share photos, thoughts, mindsets, reading news, chat with friends, share ideas and sentiments;
- Online reviews: In these platforms people can review or classify brands, services, business, and products, just to mention some examples. The reviews can add value as well as spread negative feedback. We just need to think in websites like Yelp, Zomato or TripAdvisor (Thompson, 2017);
- Social publishing: This is a way to create social media online and have feedback about it. The readers or the followers can ask questions and the author answer them. In this case we have WordPress, Tumblr or Medium (Thompson, 2017);
- Interest-based networks: This social network is built around the same interests. The audience of the social network moves around threads where the interests are the same and people search the same things. In these social media networks, we have Goodreads, Houzz, Last.fm (Afetian, 2013);
- Microblogging: This phenomenon is recent and it is a related parent of texting. Usually, with a certain amount of characters, normally 160, we can create posts through mobile devices and upload this information's, thoughts or even mood changes. In some recent world conflicts these types of media were very important to show to the world what was happening inside borders (Thompson, 2017);
- Search Engines: This is a less known social media type which includes for example Autonomy. In the past, we typed what we wanted to search and the matching results

would appear. Nowadays, we can even customize the searches, rate the results they give us and even store the search results given to us (Thompson, 2017).

Before this Social Media Age find information about companies, products or services was a hard task due to the lack of tools (Hu and Wei, 2013). With SM and the rise of Web 2.0 is now possible the engagement between companies and consumers. Therefore, brands and costumers (real and future) can create a two-way direction speech and interact in ways which were not possible before (Sigala *et al.*, 2012).

SM gave consumers / fans / followers another option to choose what products to buy, services to contract, brands products to choose in a more autonomous way. They can even learn or teach with personal experiences, experiences that they can share in SMP (Varkaris and Neuhofer, 2017).

Sharing experiences in SM platforms is called electronic word-of-mouth (eWOM) (Litvin *et al.*, 2008). When a brand uses SMA the use of the analytics will be transversal to the whole brand, as the information could be used for multiple purposes (Culnan *et al.*, 2010). Table 2 shows how different activities in the brand can be supported when using SM.

Table 2: How virtual customer environments create value

Activity Supported	Source of Value
Branding (advertising, public relations, content delivery)	Drive traffic, viral marketing, customer loyalty and retention
Sales (includes "call for action"—e.g., link to purchase item)	Revenue
Customer service and support	Cost savings, revenue, customer satisfaction
Product development	Revenue

Source: Culnan et al. (201 : p. 245)

Adopting performance indicators using structured and unstructured data collected and analyzed in SMP and making it available in meaningful and easy to read data, can benefit different areas like financial, organizational, personnel or even in systems (Rappaport, 2011).

Analyzing in an effective way the sales increase or decrease from a SM approach could be a measure for SM. Also measurable are cost reductions achieved when changing customer service from the phone to social media, or even measuring consumer satisfaction. All this information can be obtained from SMP (Ray, 2010).

In other words, we are saying that we can gain a meaningful understanding of costumer's feelings in a way that different areas of the company can take advantage of. For instance, (Constantinides et al., 2009; Turban *et al*., 2011; Etlinger *et al*., 2013; Ruhi, 2014; Prena, 2016):

- Customer Service: Sentiment Analysis can give useful information about the preferences of current and new consumers. Knowing opinions, likes, dislikes the company can improve strategies by creating or improving methods to approach to their consumers in better way in order to create a positive sentiment and erase the negative ones from them. As it is totally online, totally accessible 24/7, we are always aware of what is happening and what is being said about the brand, and so the brand can intervene instantly in the conversation to soften the sentiment if it is negative;
- Revitalize Brand: The feelings about the brand can be quantified through Sentiment Analysis. We can quantify the sentiments with the set of information we gather from the own brand, products and services provided, i.e., we can track all spikes in the comments. This analysis can be used for better marketing strategies or even for the product managers to verify their acceptance or not;
- About Competition: Why not understand what people feels about your competitors? This analysis can be useful to measure and understand what consumers feels about the competition. We can even try to predict trends and develop strategies to gain advantage over them, or even benchmark the brand process against the competitor;
- Gain Business Intelligence: The sets of information gathered via Sentiment Analysis can be provided and added to organization's Business Intelligence. Using this information in a correct manner can be precious to organizations and can help making impact decisions in order to gain advantages and leverage the business. Being in real-time means that there is a quick identification of either the positive or the negative points and take action at the moment it is needed. Even shareholders can have reports or notifications if or why the sentiments or feelings are positive or negative (He *et al*., 2015).

Drilling down and focusing in retail market to do the analysis, in this case Sentiment Analysis, we can identify some key areas where we can gain benefits when using social media analytics and therefore where we can focus social media KPI's. These areas are: Engagement, Reach, Leads and Conversions.

Engagement

When we speak about engagement we are talking about the number of likes, shares, comments, and reactions. If the organization has a low engagement it can be considered an alarm. Popularity and quality is the measure used by SMP to see the level of engagement. For example, how many people see the post of the company or how many people comment a photo or video. Typically, we can measure (Sterne, 2010):

- Clicks: The number of clicks must be coincident with the number of reactions. If we have a high number of clicks but a low number of reactions, should alert us about the type of the posts;
- Reactions: Having a high number of positive reactions would help to signal to search algorithms that the post is very popular among the SMP;
- Shares: This action indicates a conscious decision. Sharing a content means that the content is being recommended to other people that belongs to our social media platform;
- Comments: Generate comments is always good. These comments can be praise, criticism or complaints. But the important thing is to have comments as with them organizations can improve, help fans / followers or even can signal that marketing is being correctly approached and it is hitting the correct target;
- Brand Mentions: Being tagged means people are speaking about the company. So, this indicator can help understand the relevance the brand has in the market position;
- Visits: With the growth of profile visitors we can measure if the brand is getting the attention of the market. Even if there is no intention to buy, it indicates that the interest about the brand is growing;
- Active followers: These followers are only considered if they have interacted in the brand's social media platform in the past 30 days. Most of the followers just made like or follow to a page and do not visit or interact again with the brand.

Reach

Reach is an old concept and gives an estimation of how many people can really see the posts in the SMP. While engagement gives us a correct value, reach only give us an estimation because it measures the potential number of views. To quantify this point, we can measure (Fisher, 2009):

- Followers / fans: The amount of people who can effectively see the posts and said I want to follow or receive notifications;
- Impressions: Measures how many times the post was in someone's feed, even because they follow the brand / company or someone who belongs to their network shared it. This measure does not indicate how many times the post was viewed but someone received the notification of the post in their feed and could have a chance to see it;
- Traffic Data: This indicator shows us how the percentage of traffic that will arrive to the brand / company website from social media platforms. This must be a high number as that indicates we are getting in the correct way to consumers.

Lead

This measure indicates us how many engaged followers / fans are really interested to buy, or bought a product / service in the company's website. In other words, lead measures the direct or indirect impact of the SM marketing in the revenue (Henneberry, 2014). So, if we are not getting the desired revenue, either we are in the wrong SMP, we are not getting the correct audience, or our way to make the posts is not the correct one.

Customers

The goal from every brand / company is to make money with their products or services. For that they need customers, and nothing better to keep them engaged than using SMP. Keeping them engaged and interested in the brand/company products/services is a sign that they will buy again or even recommend those to their social media network. So, we must measure the number of customers who are actual buyers and the number that are really interested to buy. Measuring these values is an important part of the return on investment (ROI) coming from social media (Sukhraj, 2017).

Some of these indicators can measure how the brand/company is being seen by people, although we need to see the big picture to improve social media marketing and use this improvement to grow our revenue.

2.2 SOCIAL MEDIA IN RETAIL

The question always raised is why use SM in retail and how to take advantage from the SM. First, it is important to understand what people search and their actions in SM. People often leverage SM to identify the best products and services and to complain about products and services that are below expectations (Constantinides *et al*., 2009; Sorescu *et al*., 2011; Stephens, 2013; Niemeier *et al*., 2013; Retail, 2015).

The explosive increase of SM data creates a major opportunity for companies to leverage data analytics solutions to harness customer perceptions and better understand what people are saying on a topic, a product or about a company (Stieglitz and Dang-Xuan, 2013).

Data from SM can be used to create sets of data with more than a purpose and analyzed by different organization departments. A study by Sadler and Evans (2016) was made based in three major organizations - Tesco, Amazon and Wal-Mart - and how and for what these three organizations use SM, as we can see in the Table 3.

The authors analyzed the different SMP used by each organization, the reach values, the different accounts (by business) and the purpose of each of them. As objectives, authors identified internal communications (IC), external communications (EC) and knowledge management (KM).

For these companies, the activity in the SMP is very high, as well is high the number of interactions made in SMP by fans / followers. Thus, SM can be a great tool for retailers, having different benefits to their customers or followers (Ananda *et al*., 2015):

- Take advantage of promotions and discounts;
- Latest information on products information;
- Customer service;
- Entertaining content;
- Offer feedback.

Table 3: Tesco, Amazon and Wal-Mart: social media activity

Company	Social Media Platform	Reach	Accounts	Purpose: IC / EC / KM
Tesco	Facebook	1. 1.9 million likes 2. 70,000 likes 3. 1.8 million likes	1. Tesco 2. Tesco Mobile 3. Tesco Lotus	1. Recipes/Promotions/Customer Support 2. Promotions/Info/Entertainment 3. Promotions/Customer Support
	Twitter	1. 382,000 followers 2. 34,800 followers 3. 61,800 followers 4. 80,800 followers 5. 559,400 followers	1. @Tesco 2. @TescoNews 3. @TescoFood 4. @TescoOffers 5. @TescoMobile	1. Recipes/Competitions/Promotions/Customer Service 2. News about Tesco 3. Tips/Recipes/Competitions 4. Deals and offers 5. Promotions/Info/Entertainment
	LinkedIn	1. 170,000 followers	1. Tesco PLC	1. Recipes/ News/ Customer Support/ Company News / Promotions
	Inform App	1. 500,000 employees	1. Tesco Inform App	1. Support staff with stock levels, product information and general information
	Yammer	1. Restricted Access	1. Tesco Yammer	1. Connect colleagues and helps them work together
	Pinterest	1. 40,200 followers	1. Tesco	1. Tips/Recipes/Promotions
	Youtube	1. 6,447 subscribers 2. 2,679 subscribers 3. 32,373 subscribers 4. 4,468 subscribers 5. 14,4881 subscribers	1. Tesco 2. Tesco PLC 3. Tescofoodandwine 4. Tesco Technology 5. Tesco Lifestyle	1. Offers/News/Recipes 2. Entertainment/News/Company News/Adverts 3. Recipes and Tips 4. User guides and reviews 5. Life-hacks/Smart how-toe and tutorials
Amazon	Facebook	1. 26.5 million likes 2. 5.5 million likes 3. 5 million likes 4. 3 million likes	1. Amazon.com 2. Amazon.co.uk 3. Amazon India 4. Amazon Fashion	1. Customer Support/ News/ Offers/ Stories/ Videos/ Info 2. Customer Support/ News/ Offers/ Stories/ Videos/ Info - but UK orientated 3. Customer Support/ News/ Offers/ Stories/ Videos/ Info - India Focus 4. Offers/ Promotions/ Guides/ Suggestions
	Twitter	1. 2.7 million followers 2. 40,000 followers 3. 1.6 million followers 4. 534,000 followers	1. @amazon 2. @amazonhelp 3. @amazonuk 4. @amazondeals	1. Promotions/Videos/News/Guides 2. Customer Support 3. Promotions/Videos/News/Guides - UK orientated 4. Daily Offers
	Pinterest	1. 27,700 followers 2. 9,700 followers	1. Amazon 2. Amazon fashion	1. Themed lists of products 2. Categorised fashion lists
	Youtube	1. 71,758 subscribers 2. 55,727 subscribers 3. 1,702 subscribers	1. Amazon 2. AmazonWebServices 3. Amazon.co.uk	1. Recipes/Guides/Tutorials/Promotions/Adverts 2. Video tutorials to support their Web Services 3. Recipes/Guides/Tutorials/Promotions/Adverts
Wal-Mart	Facebook	1. 32.7 million likes 2. 1 million likes 3. 570,000 likes	1. Wal-Mart 2. Wal-Mart Canada 3. Wal-Mart Brazil	1. Recipes/Promotions/Customer Support 2. Recipes/Promotions/Customer Support 3. Recipes/Promotions/Customer Support
	Twitter	1. 737,000 followers 2. 24,500 followers 3. 236,000 followers 4. 16,400 followers 5. 16,300 followers 6. 22,100 followers	1. @walmart 2. @walmartcareers 3. @walmarttoday 4. @walmartgiving 5. @walmartlabs 6. @walmartnewsroom	1. Tips, solutions & limited time specials/ Customer service 2. Job tips and insights as well as opportunities 3. Stories and small moments – videos 4. News about their charitable doings 5. Big Data + Social + Mobile = @WalmartLabs. Powering Walmart eCommerce, we build #opensource and #cloud technology in a #devops culture. 6. Corporate announcements
	Pinterest	1. 64,200 followers	1. Walmart	1. Categorised and themed lists of items and promotions
	Youtube	1. 105,612 subscribers 2. 10,991 subscribers 3. 1,326 subscribers	1. Walmart 2. Walmart Corporate 3. Walmart Community	1. Money-saving and time-saving tips, check out the hottest new products, find awesome entertaining ideas and inspiration, and be the first to get your hands on our Rollbacks and Special Buys 2. Corporate news and activities 3. Connecting staff and customers
	Blogging	1. 20 bloggers	1. Wal-Mart moms	1. Connecting mums to share stories and offer promotions too.
	My Wal-Mart	1. Restricted Access	1. Mywalmart.com	1. For staff to keep track of their hours and benefits as well as connect with other associates.

Source: Sadler and Evans (2016, p. 4)

When developing a SM strategy, companies need to have very clear to be focused in all means to the fans / followers or future customers. The SMP are very mature and evolved so much, as they are not only an advertising channel anymore. The traditional ways of shopping are changing, as the consumers that make their buys online are growing. The increasing of competition and the growth of lack of patient of consumers is helping the increase of online shopping (Ahmad *et al*., 2015).

Moreover, brands are changing direction now; they are leveraging SM so this new direction can drive to engagement, timely conversations and personalized customer interactions (Neti, 2011). Even videos and pictures are being used as more visual elements, so brands can try a bigger engagement with consumers. Brand pages in SMP give access to their online stores, allowing to purchase a product with only a click. Companies who want to use SM data must convert unstructured data into relevant data. In this way, they can use it in supporting business strategies.

We can verify Portuguese statistics from December of 2016 (Socialbakers, 2017). In Table 4, we can compare three major SMP and their usage by retailers in Portugal. Facebook is the SMP with the higher average number of fans. Moreover, all the SMP have a relative growth of fans in Portugal.

Table 4: Average number of fans / followers in retailer's SMP, in Portugal (2016)

Social Media Platform	Number of Fans	Relative Growth
Facebook	548229	↑
Twitter	15412	↑
YouTube	5854	↑

Source: Socialbakers (2017)

Table 5 shows us the relative growth of the number of posts being made by retailers in their SMP. As we can verify, in all SMP the publications of the retailers are decreasing, and may indicate that the retailers are not so active in the SMP.

Table 5: Average number of admin posts made by retailers in SMP, in Portugal (2016)

Social Media Platform	Number of Posts	Relative Growth
Facebook	138	↓
Twitter	67	↓
YouTube	3	↓

Source: Socialbakers (2017)

Although, the average number of interactions in Facebook page of retailers is increasing, but this evidence is only in Facebook as in Tweeter and YouTube the interactions decreased (Table 6). By interactions we consider actions as like, poking, commenting, tagging, messaging, changing status, accepting invites, sending invites, checking-in or watch videos (Sarkisova, 2013).

Table 6: Average number of interactions in retailer's SMP, in Portugal (2016)

Social Media Platform	Number of interactions	Relative Growth
Facebook	140366	↑
Twitter	363	↓
YouTube	287	↓

Source: Socialbakers (2017)

In fact, the response rate in retailers Facebook increased too. However in Twitter did not happened in the same way (Table 7). In Twitter the response has decreased. For response rate we consider the percentage of new messages received that have an answer in the same day (Facebook, 2017).

Table 7: Average response rate in retailer's SMP, in Portugal (2016)

Social Media Platform	Response rate	Relative Growth
Facebook	64%	↑
Twitter	23%	↓
YouTube	-----------------	-----------------

Source: Socialbakers (2017

Table 8 indicates the type of interactions that fans / followers usually do in retailers Facebook page. The action with more percentage of interactions in retailers Facebook page is, by far, reactions. For reactions, we consider the action of reacting to a post with a certain Emoji (Bell, 2015).

Table 8: Interaction types on Facebook in retailer's page, in Portugal (2016)

Facebook Action	Percentage of interactions
Reactions	94%
Shares	3%
Comments	3%

Source: Socialbakers (2017)

Another statistic gives us the post type that retailers make in their Facebook page. Analyzing this information, we can verify that the post type with the most percentage of interactions is a post with a photo associated (Table 9).

Table 9: Facebook post types made by retailers in their Facebook page, in Portugal (2016)

Facebook Post Type	Percentage of interactions
Photo	56%
Link	35%
Video	9%
Status	0%

Source: Socialbakers (2017)

Analyzing now the other big SMP, Twitter, the likes wins for a big margin as interaction type. But comparing with Facebook, this value is lower (Table 10).

Table 10: Interaction types on retailers Twitter, in Portugal (2016)

Twitter Interaction Types	Percentage of interactions
Likes	65%
Retweets	23%
Replies	12%

Source: Socialbakers (2017)

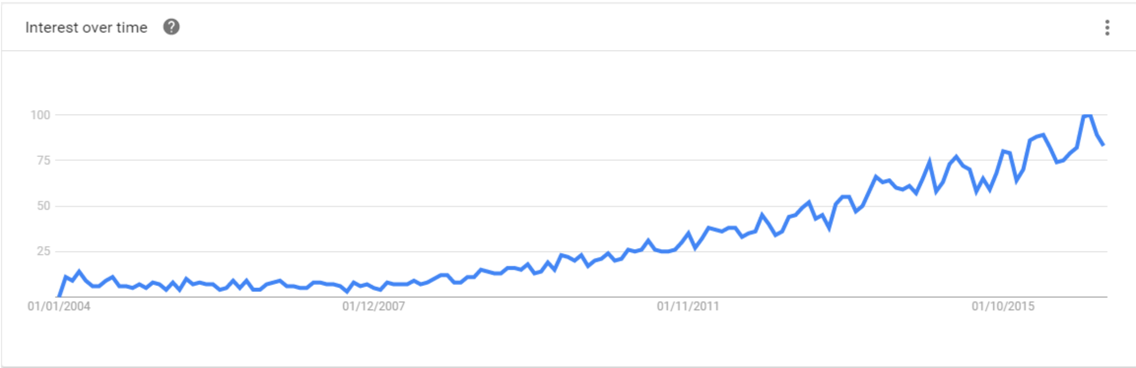
2.3 SENTIMENT ANALYSIS

SA is the process of collecting opinions, emotions or even attitudes from a text analysis. Usually it can be called "Opinion Mining" too (He *et al*., 2015).

In the early stage of SA, the methods used to analyze SA were totally focused on paper or analyzing what we call today word to mouth. With the increase of social networks, the media starts evolving the public, and the public opinion started to gain expression. So, we can now analyze and measure the sentiments of the public (Mäntylä *et al*., 2016).

If we search in Google Trends we can verify that SA only in the past few years started to have major importance and being discussed. As we can see in Figure 3, Courtesy of Google, after some fluctuation, only in the year of 2010 SA started to have major interest. In order to clarify, the numbers represent the search interest relative to the highest point of the chart for the specified region and time interval. A value of 100 is the peak in popularity of the term. A value of 50 means that the term had half the popularity. Likewise, a score of 0 means that term has less than 1% of popularity relative to the highest point.

Figure 3: Sentiment Analysis topic interest over time



Source: Google Trends (2017)

The virtual interaction between people with the same interests is growing, as are growing the SMP and the frequency people write in SM their thoughts and point of view about brands, products and services (Asur and Huberman, 2010). Companies need to have an understanding about their own consumers. Having these understanding, they will know what they like, dislike, what they want or they do not want (Fan *et al*., 2014). When consumers write comments in a SMP, they usually express their opinions in words about what they are feeling at the moment they are writing the comment: joy, sadness, happiness, unhappiness, anger, and other human feelings (Munezero *et al*., 2014). These interactions can show the feelings of consumers and can give a set of important information (Liu *et al*., 2007).

The main purpose of SA is to identify statements, opinions (positives or negatives) and reactions in online content created by users in SMP (Xiang *et al*.,2015:121). This type of analysis can be made using posts and comments in Facebook, tweets in Twitter, emails, blogs. Moreover, SA allows, for instance, to answer the following questions:

- Is the product review positive or negative?
- Is the customer satisfied or not?

- Based on comments, how are people responding to a determined campaign/product release?

In the Retail case, we use the SA in order to gain a meaningful understanding of the feelings of fans / followers (Culnan *et al*., 2010).

Most of the software used to make this analysis uses Natural Language Processing techniques as well as a number of complex rules-based algorithms, in order to have an exhaustive and detailed set of information about the feelings that had been expressed in words in SMP (D'Andrea *et al*., 2015). The analyzed sentiments can be classified as: positive, negative or neutral (He *et al*., 2015). To get these sentiments the process uses text analytics, linguistic and accepted language processing so it can verify and analyze only the subjective information.

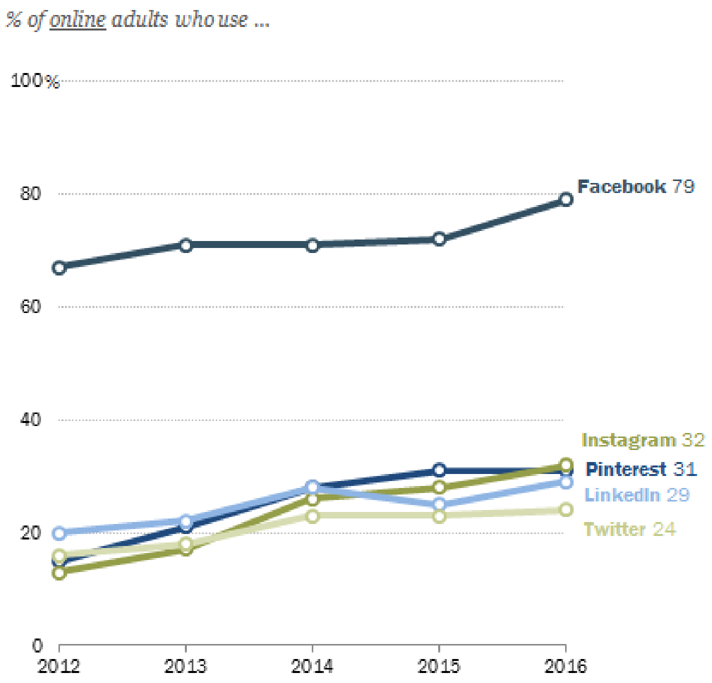
Before performing a SA, retailers must know how can SA help to improve the business. He et al. (2015) explicit the following ways:

- Customer Service: Nowadays brands customers (happy and upset ones) go straight to the internet. Customer service must shift to a proactive activity. So, customer service can give useful information about the preferences of current and future consumers. Customer service can prioritize service or response times to customers prioritizing first the most unsatisfied (Mandelbaum, A., 2014);
- Revitalize Brand: The feelings about the brand can be quantified through SA. We can quantify the sentiments with the set of information we gather from the own brand, products and services provided, i.e., we can track all spikes in the comments. The extraction and analysis of all this information can be aggregated and quantified in a meaningful way for decision makers;
- About Competition: Why not understand what people feel about your competitors? This analysis can be useful to measure and understand what consumers feel about the competition. We can even try to predict trends and develop strategies to gain advantage over them. Or even benchmark the brand process against the competitor;
- Gain Business Intelligence: The sets of information gathered via SA can be provided and added to organizations BI. Using SA metrics in a correct manner, can be precious to organizations and can help making impact decisions in order to gain advantages and leverage the business. Being in real-time, we can quickly identify either the positive or the negative points and take action at the moment. Even the shareholders can have reports or notifications if or why the sentiments or feelings are positive or negative.

THE IMPORTANCE OF FACEBOOK?

Facebook is one of the best places where we can find sentiments. Nowadays, any event raises several reactions and comments. Facebook is also the most popular SMP as we can see in the United States of America (Figure 4) and also in Portugal (Statcounter).

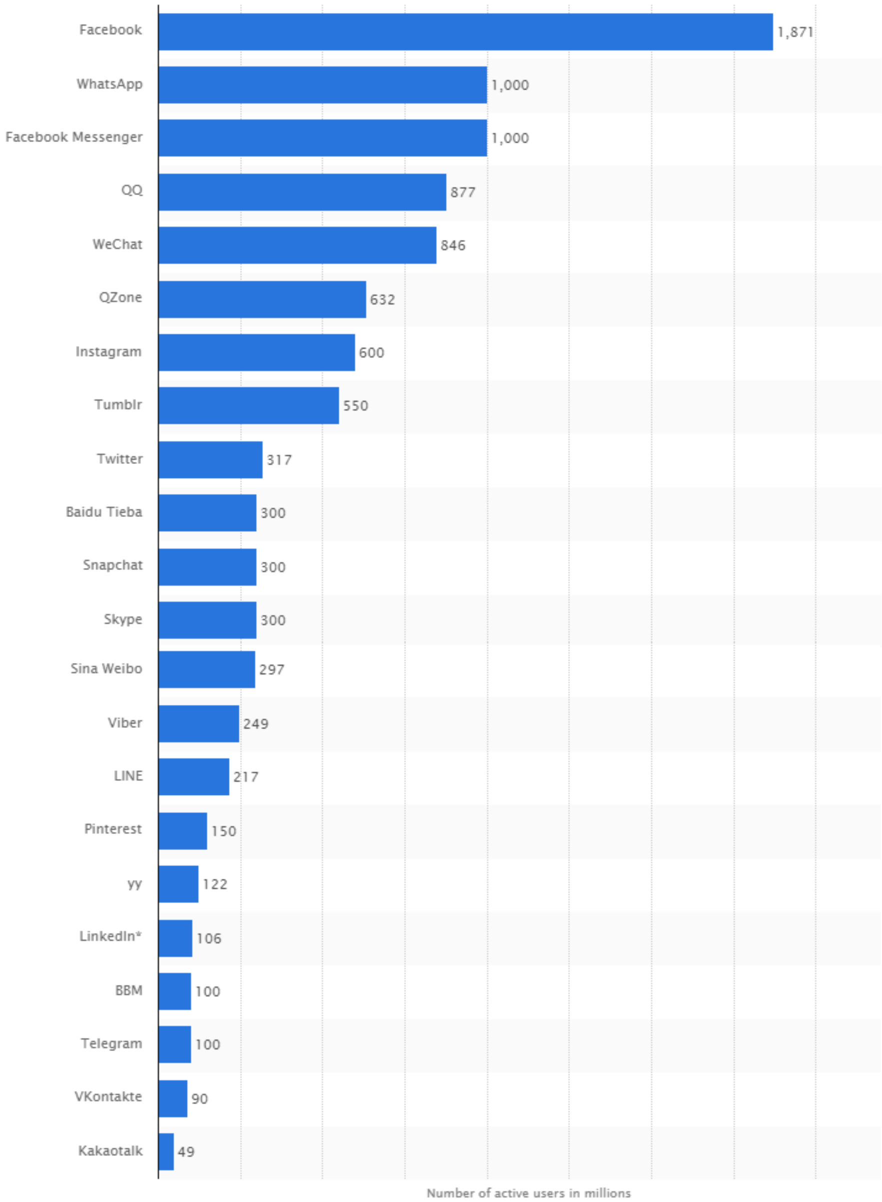
Figure 4: Most popular social networks



Source: Chaffey (2017)

Facebook has, worldwide, over 1.94 billion monthly active users and there are 1.15 billion mobile daily active users (Facebook, 2017). Figure 5 shows the number of active users by SMP. We can have people who may have accounts in one or more SMP, but still, Facebook has the higher number of active users regarding the other SMP.

Figure 5: Social media platform with the higher number of active users worldwide



Source: Statista (2017)

Facebook has more advantages over the other SMP's, such as Instagram, Youtube or Pintarest. Another advantage of Facebook is related with the character limitation (does not have) and its worldwide use. In this way, Facebook gives more type of interactions to analyze, and so, the results will be more accurate. These contents shared by brands can have different interactions such reactions, shares or comments. Globally we can easily create communications in order to inform consumers, as the posts we share in the company page can be totally public. Fans / followers can get information about all the brand page activity, just following that. From a corporate point of view Facebook should be used to create shareholder communities as it is made easy to manage the structural and flow properties in a relationship (Kiezmann et al., 2011).

Table 11 shows the terminology used in Facebook for user activities. This table shows terminology of engagement and how fans / followers can interact with brands Facebook page.

Table 11: Facebook terminology

User	Activity	Facebook Term
User 1	Status update	Post, Status; Status post
User 2	Agree with status update of user 1	Like; Love, Haha; Wow
User 2	Disagree with status update of user 1	Sad; Angry
User 2	Make a comment	Comment
User 2	Reply a comment	Reply
User 2	Share User 1 status update	Share

Source: Facebook (2017)

2.4 SENTIMENT ANALYSIS STUDIES IN RETAIL

SA can be used to gain contributes for company performance, the brand performance and check consumer/customer loyalty. The usage of SM by brand and the interactions made with the consumer/customer can give, for instance, the reputation of the brand. In the literature, we can find some studies about SA in retail.

2.4.1 COSCO AND WALMART

In this study, the authors Wu *et al.*, (2015) focused in two large retail chains in the world: Cosco and Walmart. Tables 12 and 13 show their findings regarding the SA in Twitter about four identical products, in the same time interval and that are sold in both retail chains. The researchers analyzed a high number of comments (229.517) and then made the "opinion mining" in order to get the overall sentiments about the four products of both retail chains. Table 12 shows the data collected in Cosco Twitter and shows the number of mentions and the sentiment that customers have about the products. Table 13 shows the data collect in Walmart Twitter and shows the number of mentions and the sentiment that customers have about four distinct products.

Table 12: Cosco Twitter data from december 1, 2014 – February 29, 2015

Product name	No. of mentions	Positive sentiment	Neutral sentiment	Negative sentiment	Overall sentiment
Muffin	1,103/229,517 (0.48%)	191/1,103 (17.32%)	850/1,103 (77.06%)	62/1,103 (5.62%)	Positive
Cookie	784/229,517 (0.34%)	113/784 (14.41%)	609/784 (77.68%)	62/784 (7.91%)	Positive
Pizza	13,242/229,517 (5.77%)	1,850/13,242 (13.97%)	10,759/13,242 (81.25%)	633/13,242 (4.78%)	Positive
Chicken	2,670/229,517 (1.16%)	424/2,670 (15.88%)	2,051/2,670 (76.82%)	195/2,670 (7.30%)	Positive

Source: Wu et al. (2015) : p. 1628

Table 13: Walmart Twitter data from december 1, 2014 – February 29, 2015

Product name	No. of mentions	Positive sentiment	Neutral sentiment	Negative sentiment	Overall sentiment
Muffin	41/246,442 (0.01%)	13/41 (31.71%)	26/41 (63.41%)	2/41 (4.88%)	Positive
Cookie	506/246,442 (0.21%)	84/506 (16.60%)	366/506 (72.33%)	56/506 (11.07%)	Neutral-positive
Pizza	410/246,442 (0.17%)	60/410 (14.63%)	310/410 (75.61%)	40/410 (9.76%)	Positive
Chicken	468/246,442 (0.19%)	72/468 (15.38%)	344/468 (73.51%)	52/468 (11.11%)	Neutral-positive

Source: Wu et al. (2015) : p. 1629

With these two tables, both companies can verify each other customers' sentiment about same products. These analyzes give them the possibility to verify the competitors' position in the market and give them the opportunity to improve either the quality of the products, either to explore new marketing methods or even to decision makers take another type of decisions in order to obtain leverage against their competitors. This is the way business goals will be as the North Star for businesses and will help decision makers formulate and prioritize their social media measurement initiatives (Etlinger *et al.*, 2013).

In the conclusion of the study the authors wrote "…results of the case study suggest a business opportunity for developing a social media data application for competitive analytics and intelligence purpose because companies in general want to know how consumers feel about their products and services and those of their competitors." (Wu *et al.*, 2015: p. 1633).

2.4.2 FACEBOOK POSTS

Another study was made using Facebook posts. "From the total of 759 shared posts, we have removed 134 posts published by the site moderator, leaving 625 posts published by the page fans. Of those 3 were removed due to the difficulty in recognizing the used language and an additional 11 after being labeled as spam, leaving 611 user posts for qualitative analysis." (Pletikosa and Michahelles, 2011: p. 177).

Table 14 reproduces the researchers' findings, showing the topics that fans wrote in the brand page.

Table 14: Post reason distribution

Topic Group	Occurrences	Percentage	z
Product	318	52%	14.539*
Sales	79	13%	3.02**
Brand	46	8%	2.306***
Competitor	26	4%	0.759
Facebook Contest	20	3%	1.452
Company	11	2%	1.883
Environment	3	0%	
TOTAL:	**489**	**80%**	

*Note: (*p<0.0001, **p<0.005, ***p<0.05)*

Source: Pletikosa and Michahelles (2011: p. 177)

The results show that there is a significant difference ($z = 14.539$, $p < 0.0001$) in the proportion of Product posts (52%) compared to the Sales (13%). Moreover, the proportion of posts regarding Sales is significantly larger ($z =3.02$, $p < 0.005$) compared to the proportion of Brand posts (8%), while the mentions of the brand is significantly larger ($z =2.306$, $p < 0.05$) compared to the number of Competitor (4%) posts.

Table 15 shows all the category of the posts, that the fans commented in the brand page.

Table 15: Post categories

Post Category	Occurrences	Percentage	z
Suggestions & Requests	170	28%	-0.001
Affect Expression	169	28%	0.195
Sharing	165	27%	4.592*
Information Inquiry	98	16%	7.09*
Complaints & Criticism	23	4%	-0.003
Gratitude	22	4%	3.111*
Praise	5	1%	
Competitor Reference	22	4%	
TOTAL:	674	N/A	

*Note: (*p<0.0001)*

Source: Pletikosa and Michahelles (2011, p. 178)

Comparing the values of Information Inquiry (16%), Sharing has a significantly larger number ($z= 4.592$, $p < 0.0001$). Moreover, the proportion of Information Inquiry (16%) is significantly larger ($z = 7.09$, $p <0.0001$) compared to Complaints & Criticism (4%). The proportion of Gratitude (4%) is significantly larger ($z = 3.111$, $p < 0.0001$) compared to the proportion of least occurring post category, i.e., Praise (1%). For the remaining post categories, no significant difference was found ($p > 0.05$).

As conclusions, the researchers presented us with the following (Pletikosa and Michahelles, 2011: p. 180):" Facebook brand pages support the social media marketing opportunities and goals for building brand awareness, gathering insights and knowledge for future steps, community involvement and engaging in open and honest dialog, as presented in the related work section. Marketing practitioners could use the topic-category frequency of occurrence as a measure for successful social media marketing utilization over time.".

2.4.3 BURBERRY CASE

In 2011, it was made a study about a well-known brand: Burberry (Phan *et al.*, 2011). Burberry was in 2011, one of the most active luxury brands. It had over 7 million fans on Facebook (as compared to about 5 million fans for Gucci and 3 million fans for Louis Vuitton) and over 400.000 followers on Twitter (as compared to 100.000 for Gucci and over 200.000 for

Louis Vuitton). These values were from 2011. The brand had some hard times when it was associated to hooligans, and even more when a group of them called themselves the Burberry Boys. Rose Marie Bravo, CEO of the brand, started the brand revitalization. One of the strategies adopted to revitalize the band was using a SM strategy. In 2008, Burberry spent 1,5% of its total advertising expenditure and in 2010 spent 40% of total advertising expenditure. Despite the use of the social network platforms and the use of the standard actions, Burberry launched in October of 2009, their own social network site and they called it Art of the Trench. The main goal was stimulating the interaction between consumers and brand's culture.

The next step was to start selling Burberry products directly in Facebook. More connections were created between consumers and the brand and the engagement grew up. As Burberry, many brands are focusing in SM marketing and especially in a customer-focused strategy. First must come brand strategy and then SM. In 2011, Burberry was considered the best-ranked luxury brand by Famecount (Phan *et al*., 2011: p. 213). The study results showed that 15% of fans of the Facebook page maintained regular interactions with the brand. As this amount of interactions was considered low, the researchers proposed three actions. The first one, was creating a SMP where users could discuss the brand and their products. This definition commonly is the definition of a forum. The second one was involving consumers/fans in some decision-making processes or participating in products creation. The third one was electing the "best" fan to transmit Burberry values.

Moreover, researchers concluded that: i) the image of the brand is not created only by marketers but it has influence of consumers and user communities; ii) the brand image requires engagement in SM and interaction between the brand and consumers; and iii) the SM strategy approach the brand with young consumers and always control what contents are being shared in SMP, as the image of the brand can be seriously affected by a bad choice of contents to share in a SMP.

2.4.4 LOUIS VUITTON CASE

In the Louis Vuitton study, the focus was the well-known brand (Kim and Ko, 2010). The study started with the distribution of questionnaires to collect data. This questionnaire was restricted only to consumers who bought luxury products in the previous two years and focused in five properties of the SM: entertainment, customization, interaction, word of mouth, and trend.

The study proved the effectiveness of SM in luxury brands, in particular on customer relationships and purchase intention. A big number of brands have been entering in the luxury market, because luxury fashion is considered a high value-added and of course with very high profit margins. As the lower sectors started to be competition to the luxury brands, one of the problems is how to secure their customers (Kim and Ko, 2010).

The study focused in the next key points to evaluate brand performance when using SM:

- Intimacy: Refers to feelings of closeness, connectedness, and Bondedness;
- Trust is considered one of the keys to an enhanced relationship between a consumer and a certain brand, and it has been recognized recently as a core variable of long-term relationships with customers;
- Purchase intention: Is defined as the consumer's possibility of purchasing in the future.

The research aims to understand how SM affected customer relationships (i.e., intimacy and trust) and purchase intention. The findings show that SM has an important role for luxury brands. Intimacy has a significant positive influence coming from entertainment and word of mouth and in turn had impact on customer relationships and purchase intention. Moreover, entertainment, customization and trend have influence in trust, while purchase intention is influenced by entertainment, interaction, and word of mouth.

The study concluded that: i) in the present competitor environment, future behavior of consumers is a key strategic asset and must be monitored; ii) trust gain strength with interactions between consumers in SMP; and iii) the usage of SM appears to be appropriate for retaining old customers and attracting cross-shoppers. So, managing customers have increasing importance as competition is more intense. Moreover, luxury brands must engage in SM to anticipate contributions to the brands by costumers providing in this way new luxury values to them (Kim and Ko, 2010).

2.5 BUSINESS INTELLIGENCE

BI is a set of practices, applications and technologies whose purpose is to collect, integrate, analyze and present business data. The goal of BI is to better support business decision making. We can say that BI systems are data-driven Decision Support Systems (Watson and Wixom, 2007).

It is anticipated that BI could provide unique data collection and analytical research, as well as development opportunities using data and text mining techniques (which is the process of acquiring appropriate knowledge from structured and unstructured data (Turban, Sharda, & Delen, 2011).

Zeng *et al*. (2010: p. 15) distinguished the concepts of SMA and SM intelligence: "*Social media intelligence aims to derive actionable information from social media in context rich application settings, develop corresponding decision-making or decision-aiding frameworks, and provide architectural designs and solution frameworks for existing and new applications (...)*."

BI has the goal to support the organization decisions by using important analytical data. SM has its sources in personal knowledge, opinions and attitudes. Merging both we have the concept of Social BI (Lovett, 2011).

2.5.1 SOCIAL BUSINESS INTELLIGENCE

Social BI is a new trend and is emerging in BI. There is no Social BI definition. Many authors refer to this topic as social analysis, BI 2.0, SMA, social intelligence, SM intelligence or even social BI (Dinter and Lorenz, 2012).

Social BI has its focus in the analysis of social media in order to quickly and easily make decisions and create actions (Gilasgar *et al*., 2016). The improvements in the technology area changed the process how fans / followers interact with the brands. Nowadays, most of the consumers interact immediately with brands via SMP (Gallinucci *et al*., 2013). This change of habits and the chance to get feedback or reactions from all over the world explain why social BI is a requirement to decision makers. The multiple ways to collect the information and the chance to collect data from all over the world fans / followers proved to be valuable and provide a new set of opportunities (Roblek *et al*., 2013). For better understanding, we will call to this topic social media intelligence when we refer to the social media data integration into the organization BI (Berlanga *et al*., 2015).

Thus, social BI supports a broad range of processes in research and development, sales, customer service, and operations, just to name a few (Bose 2011). One of the goals of the project is to incorporate all the data gathered in the organization BI, to have a more clear understanding about the sentiment that fans / followers have of the brand.

The huge growth of the data produced by SM can be confirmed day by day, even can say that is tangible (He *et al*., 2015). If we analyze data we can verify that the number of fans / followers and potential customers that are interacting with each other through the SM, is increasing (Patroni *et al*., 2016). Another perception we have is the nature of the media is expanding too. SM nowadays can have a form of posts, tweets, images, videos, gifs, geographic data. The online platforms and social media are gaining each day more preference from the fans / followers as a means to interact with brands, creating a huge opportunity to social BI to emerge (Roblek *et al*., 2013).

Social BI can give us tools that can show us the costumer engagement with the brand and help to predict trends or even guide them (Kim and Ko, 2010). Percept the sentiments of fans / followers and understand them it is another tool that social BI gives us (Ananda *et al*., 2015).

The increased focus in social BI reveals how important is SM, being seen by business and fans / followers and how their interaction is changing. The high availability of and accessibility of the SM data changed the focus of the organizations (Berlanga *et al*., 2015). Now they are focused on customer engagement. Recent research on SMA has emphasized the need to adopt a BI based approach to collecting, analyzing and interpreting SM data (Murdough, 2009; Heijnen *et al*., 2013;).

2.5.2 KPI

A key performance indicator (kpi) measures the effectiveness of a company achieving the key business objectives (Investopedia, 2017). The use of kpi helps to understand how the department or organization performance and should point us the right direction und understand if we are in the right path to achieve our strategic goals (Konsta and Plomaritou, 2012). With the integration on the data collected in SMP into the organizational BI we need to have the relevant kpi. These indicators can be used by different organization areas, and they will have the same data source (Pentina and Koh, 2012). In SM, the kpi's are not the ones we generally use, they must reflect SM measures.

From SM, we can have as main indicators (Podobnik, 2013):

- Number of likes: quantity of likes made by fans/followers in a post;
- Number of shares: quantity of fans/followers that shared a post;
- Comments: number of comments made by fans/followers in a certain post;
- Sentiments: sentiments of the fans/followers collected via SA;
- Number of followers: number of persons who made the action of following the page.

But also, we can combine some metrics and then we will have metrics such as (Fontein, 2016):

- Number of mentions: a mention happens when we tag another person in a comment of the brand page;
- Referral traffic: is the amount of traffic generated in the website of the brand that comes from SMP;
- Share of voice: measures the number of mentions that a brand receives in relation to competition;
- Net promoter: amount of people that recommends the brand in SMP;
- Video views: the amount of people who watched the video till the end;
- Conversions: measures the positive actions of a visitor of the website. Usually we can measure if the brand is converting the visitors;
- Sales revenue: number of consumers/buyers who came from a SMP;
- Issues resolved: measures the efficiency of the brand solving consumers issues;
- Customer lifetime value: after earning the trust and loyalty of a consumer, this indicator measures the revenue that a consumer will generate.

These types of metrics can help the responsible from different departments of the organization to make decisions (Sejati *et al.*, 2015). The big issue is choosing the correct kpi to monitor, so we can be aligned with our goals. So, in order to accomplish that, we need to understand very well our purpose and what we really need and want to measure, as the correct kpi will give us the benefit and will not drive us to an incorrect path. As Ruhi (2014: p. 5) says, "*… the relationship between corporate objectives, supporting business unit metrics or kpi's* *and social media metrics should be identifiable.".* Table 16 gives examples of kpi's, what they measure and what they are good for.

Table 16: Kpi's examples

Kpi	What they measure	What they are good for
Audience Growth	Size of Network	Increasing exposure + Reach
Engagement Rate	Comments, Replies, etc.	Building Trust
Amplification Rate	Shares/RT, etc.	Feedback of the comments in consumers
Response Rate	% addressed queries raised by customers	Building Trust
Sale	Sign up/downloads	Revenue
Assisted Social Conversion	Social touchpoint in acquisition tunnel	Indirect social impact has on acquisition of merchandise
Social Visits	Referral Traffic	Potential prospects in company domains

Source: Sukhraj, 2016

2.6 SOCIAL MEDIA METRICS

"*Metric is a quantifiable measure that is used to track and assess the status of a specific process and are important because they are comprised of a wide swathe of all trackable areas.*" **(**Taylor, 2017).

So, even for SM and in concrete for SA, there are metrics to track sentiments contained in text. The authors Wilson *et al*. (2009) and Tan *et al*. (2012) studied and developed a SA phrase-level that verifies what type of tone is used in the text. This concept, polarity or sentiment orientation, indicates if the text has a positive, negative or neutral sentiment. Identifying sentiments in sentences is a very complicated problem whose ambiguities in languages lead to the existence of different optimized algorithms used by machine learning techniques. Several sentiments algorithms were compared by Thelwall *et al*. (2010) to study the performance of the correct analysis and the correct score of positive and negative sentiments. In the research, the participants were asked to score positive and negative sentiments between [-5; +5]. The goal was to understand and categorize which words were associated with positive, neutral or negative sentiments. A common used polarity scale goes from [-1; +1]. The metrics to evaluate the performance of the brand and to give information to decision makers are provided by SMA. The most frequent and important metrics used by brands, or/and type of calculus are presented in the Table 17.

Table 17: Metrics used by brands

Performance metric (Saothawan, 2014)	Data Collected	Description	Method	Object Measured
Public agreement	Number of likes	Number of likes of a specific post in relation to average number of likes. It indicates how much a post is publicly accepted: $\frac{Number\ of\ likes}{average\ number\ of\ likes}$	Statistical method	Post
Popularity	Number of shares	Number of shares of a specific post in relation to average number of shares. It indicates the popularity: $\frac{Number\ of\ shares}{average\ number\ of\ shares}$	Statistical method	Post
Public Involvement	Number of comments	Number of comments of a specific post in relation to average number of comments. It indicates how much a post triggers user to publicly show their opinion: $\frac{Number\ of\ post\ comments}{average\ number\ of\ post\ comments}$	Statistical method	Post
Polarity level	A single comment on a status post	A user comment usually contains user sentiment on a post. Measuring is enabled by Sentiment Analysis methods.	Text Analysis	Comment
Public polarity	Overall comments on a specific status post	Measuring overall sentiment of users. For example, by the average sentiment level of all user comments	Aggregated polarity level Statistical method	Comment

Source: Saothawan (2014)

With this analysis, it is possible to have fans / follower's opinion and sentiments about the brand, using the posts timeline in brand Facebook's page. Stelzner (2013) in his study of 2013 reported that only 26% of brands measures social media activity. In other research, Stodder (2012) found that 32% of the brand who participated in a study realized by himself, did not analyze SM.

The brands, in order to obtain performance results of SM usage, usually analyze the posts made in SMP to calculate the public agreement, popularity and the public involvement.

Each brand set the timeframe to measure the performance and the percentage range too, as shown in Table 18. After doing the mathematical operations to calculate the metrics values, it is needed to evaluate the results. The percentage range must be predefined, so it can be measured in a standard way (Saothawan, 2014).

The performance metrics are not standard to all brands. Each brand has its own purpose and analysis, so the evaluation ranges must be set by business. For a general orientation can be used the values in Table 18, to get the perception of a user-generated post by business (Saothawan, 2014).

Table 18: Evaluation range examples

Public Agreement		Popularity		Public Involvement	
Metric value	Evaluation	Metric value	Evaluation	Metric value	Evaluation
0-30	Neutral	0	Neutral	0-33	Low
31-66	Slight public agreement	31-66	Slight popular	33-66	Slight public involvement
67-100	Average public agreement	67-100	Average popular	66-100	Normal public involvement
100-149	High public agreement	100-149	High popularity	100-133	High public involvement
>150	Very high public agreement	>150	Very high popularity	>134	Very high public involvement

Source: Saothawan (2014)

However, the performance metrics described to verify performance on SMP are not 100 percent accurate and reliable. If we use them to make a competition comparison we must remember that there are services that boost likes number (e.g. boostlikes.com). With this service, the information collected could have false values, and all metrics that depend on the likes number would not be accurate (Wiltshire, 2017). Although, if it is our own brand we can correct metric ranges, but if it is a competitor we must have always this issue present. The main metrics used are presented in Table 19, for page level, and in Table 20, for post level.

Table 19: Facebook insights on page level

Metric	Description
Logged-In Page Views	Total number of brand page Views.
Page Consumers	The number of unique users who clicked on any of the brand page content. It includes link clicks, photo clicks, video plays and other clicks.
Page Consumptions	Number of clicks on any of the brand page content.
Page Engaged Users	The number of unique users who engaged with the brand page, including any click or story created.
Page Impressions	Number of impressions seen of any content associated with the brand page. It can be possible to have multiple impressions of the same post. For example, a fan might see a Page update in their news feed once, and then a second time if a friend shares it.
Page Organic Impressions	Number of times posts were seen in News Feed, Ticker or on visits to the brand page.
Page Organic Reach	Number of unique users who visited the brand page, saw the page or one of its posts in News Feed or Ticker.
Page Paid Impressions	Number of impressions of a Sponsored Story or Ad pointing to the brand page.
Page Paid Reach	Number of unique users who saw a Sponsored Story or Ad pointing to the brand page.
Page Reach	The number of unique users who have seen any content associated with the brand page. The Page Reach number might be less than the impressions number since one person can see multiple impressions.
Page Stories	The number of stories created about brand page. Stories include: • Likes of the brand page; • Likes, comments on, or shares of brand page posts; • Answers to a question asked; • Responses to events; • Mentions of brand page; • Tags of brand page in a photo; • Check-ins or recommendations of brand page.
Page Talking About This	Number of unique users sharing stories about brand page.
Page Viral Impressions	Number of impressions of a story published by a fan / customer about brand page.
Page Viral Reach	Number of unique users who saw brand page or one of its posts from a story published by a fan / customer.

Source: adapted from Loomer (2012)

Facebook Insights or Facebook Analytics, is a platform for data and information that shows how the audience is interacting with your page. Facebook Insights helps brands to perceive how the content of their pages is resonating with their fans / followers or the growth of the page (Lua, 2017).

Table 20: Facebook insights on post level

Metric	**Description**
Negative Feedback	Number of times users have clicked the “x” button, clicked “hide,” clicked “hide all” or reported brand page posts as spam.
Post Consumers	Number of unique users who clicked anywhere in brand post.
Post Consumptions	Number of clicks anywhere in brand post.
Post Engaged Users	Number of unique users engaged with brand post. It includes any click or story created.
Post Impressions	The number of times a post from your Page is displayed, whether the post is clicked on or not. People may see multiple impressions of the same post. For example, a fan might see a Page update in their news feed once, and then a second time if their friend shares it.
Post Organic Impressions	The number of impressions of your post in News Feed, Ticker or on your Page’s Timeline.
Post Organic Reach	The number of unique users (Fans or non-Fans) who saw your Page post in News Feed, Ticker or on your Page’s Wall.
Post Paid Impressions	The number of impressions of your Page post in an Ad or Sponsored Story.
Post Paid Reach	The number of unique users who saw your Page post in an Ad or Sponsored Story.
Post Reach	The number of unique users who saw your Page post. The Post Reach number might be less than the impressions number since one person can see multiple impressions.
Post Stories	The number of stories generated about your Page post. Stories include: • Likes, comments on, or shares your Page post; • Answers a question you’ve asked; Responds to your event.
Post Talking About This	The number of unique users who created a story about your Page post.
Post Viral Impressions	The number of impressions of your Page post in a story generated by a friend.
Post Viral Reach	The number of unique users who saw your Page post in a story from a friend.

Source: adapted from Loomer (2012)

2.7 Brand Management

Maintaining, improving, and upholding a brand so that the name is associated with positive sentiments is part of the process of BM (Shamoon *et al*., 2012). It involves many variables, such as customer satisfaction, in-store presentation or competition. BM has its focus on the brand and directing it to fans / followers. Using a proper brand management can help the brand to achieve higher sales not only in one product, but on other products of the brand. SM is a part of the BM process (Zailskaitė-Jakštė and Kuvykaitė, 2016).

SM can be a powerful tool to construct consumers' attitudes, behaviours, intentions, increase reliability and increase consumers perception about the brand and brand products (Goldsmith and Horowitz, 2006; Bambauer-Sachse and Mangold, 2011). Moreover, brand loyalty, in SM context, can be defined as the primary choice of the consumers when it starts a purchase process. The continuous positive "word-of-mouth" or posts creation in SM networks about a brand can be included in the definition (Wang *et al*., 2012).

Authors make a differentiation between company created content and user generated content in SM. This distinction is made because company-created communication in SM is under the management of companies, while consumer-generated communication in SM is independent of the companies' control (Bruhn *et al*., 2012; Schivinski and Dąbrowski, 2013). Additionally, the communication in SM without consumers' interactions are not reliable and will not influence consumers. Interaction between the brands and consumers can strength the brand position (Beuker and Abbing, 2010).

Other authors used McQuail's four-category motivation classification (entertainment, integration and social interaction, personal identity and information) in order to apply it in SM communications (Brandtzæg and Heim, 2009; Calder *et al*., 2009).

Brand Management and Social Media are linked in the present with the growth of the Internet and Social Media Platforms usage.

BRAND ANALYSIS

Opinion Mining can be a tool to make a brand analysis. As it is essential to brands to monitor their reputation, SA can provide the tools to do it. The content generated in SM networks can provide to the brand the accurate vision that fans / followers have from the brand, products or even business decisions like promotions or price changes (Dey *et al*., 2011).

We can make another type of analysis about the brand like for example (Arboleda *et al*., 2017):

- What is the impact of the customer voice on the brand reputation?
- How will this affect our competitive advantage over competing brands?
- Will this affect our brand loyalty and positioning?
- How do we reposition to offset this negative perception?
- Are the consumers loyal to the brand?

Online content presented in social network platforms can be the catalyst for sentiment ranges, along with other behaviors. These reactions and comments online have subjective sentiments and opinions (Pang, and Lee, 2008; Bhadane *et al*., 2015).

The sales volume can be influenced by sentiments and opinions. Some brands are already realizing the importance of SA so they can get the sentiments or opinions about a certain product, service or business decision, that can give them competitive advantage, or learn about weaknesses and strengths, or even catch new opportunities to weaken competition (Wijnhoven and Bloemen, 2014; Serrano-Guerrero *et al*., 2015).

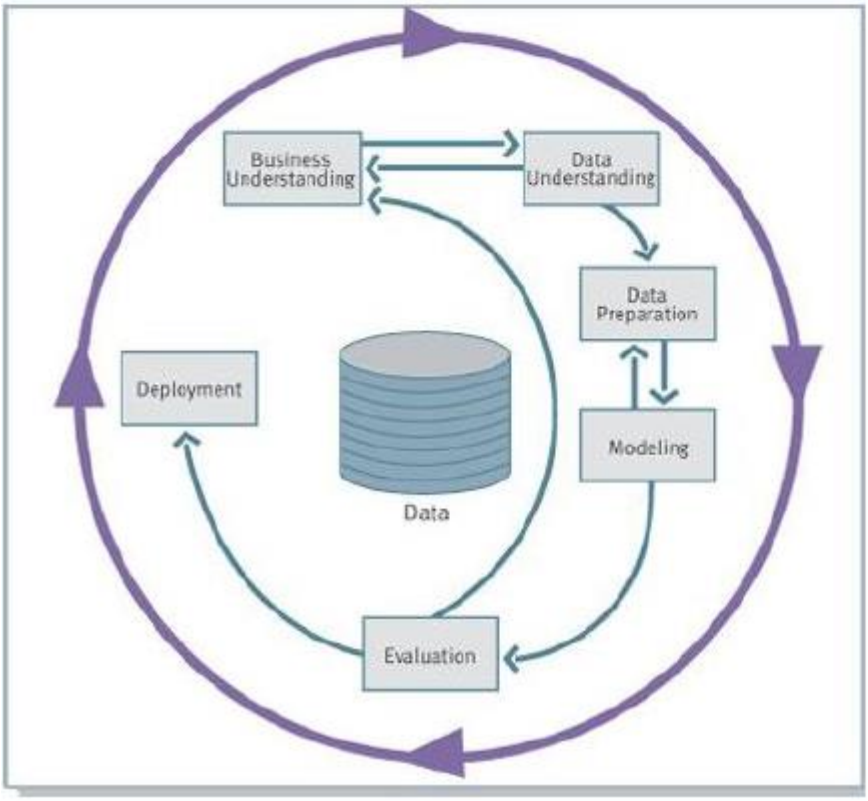
When we apprehend public opinion about a brand, SA will help us extract sentiments from thousands of online sentences made in comments over all SMP, and classify that information in a valuable manner. This requires a big effort and a lot of work in order to go from data gathering to the conclusions of the analysis (Gilasgar *et al*., 2016).

3 METHODOLOGY

According with the objectives proposed, the CRISP-DM (Cross Industry Standard Process for Data Mining) was the methodology adopted as it is a standard framework and where our practical work fits perfectly (Chapman *et al*., 2000; Nadali *et al*., 2011).

CRIPP-DM is oriented to business and this is the primary reason to choose the usage of the methodology. This methodology has six phases (Figure 6), that are business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Chapman *et al*., 2000).

Figure 6: CRISP-DM Methodology



Source: Chapman et al. (2000 :13)

Understanding the business and its problems is one of the subjects of the business understanding phase. Therefore, the formulation of the problem, determination of the plan and correspondent project and objectives must be done at this phase. Also, must be defined and clear what are the goals of the data mining. In this research, it will be mined the Facebook page of the brand so it can be extracted all posts and comments made in the page.

After this phase, it starts the data understanding phase. This phase begins with data gathering. This phase includes the collection of initial data, the description of data, the exploration of data and the verification of data quality. Associated to this phase appears the data preparation step, where we select, clean, construct, integrate and format data.

The next step is to choose and parametrize the modeling techniques. This is the modeling phase. If we need some special data requirements, as for example, the need of the comments to be all in English, we must set this requirement in the previous phase. So, in the modeling phase we have: modeling technique selection, test design generation, model's creation and model's assessment.

In the evaluation phase, we evaluate the model and its correspondent construction, so we can ensure that all steps we pass through are correct and there are no flaws in the model design. When this evaluation is assured, we can start deciding what we would do with the results we got in our experience. It is in this phase that is verified if the model is suitable with the objectives proposed in the initial phase.

In the last phase of CRISP-DM methodology the results are organized in a way that the researcher or customer can use them. This organization of the data can be as simple as a report or can be complex, as we can implement some data mining models in the organization (Chapman *et al*., 2000).

This last phase has some more steps, like for example the deployment plan, monitoring plan, the final report production and the review of the whole project. The big advantage of CRIPS-DM is that we can be able to return to earlier stages if needed.

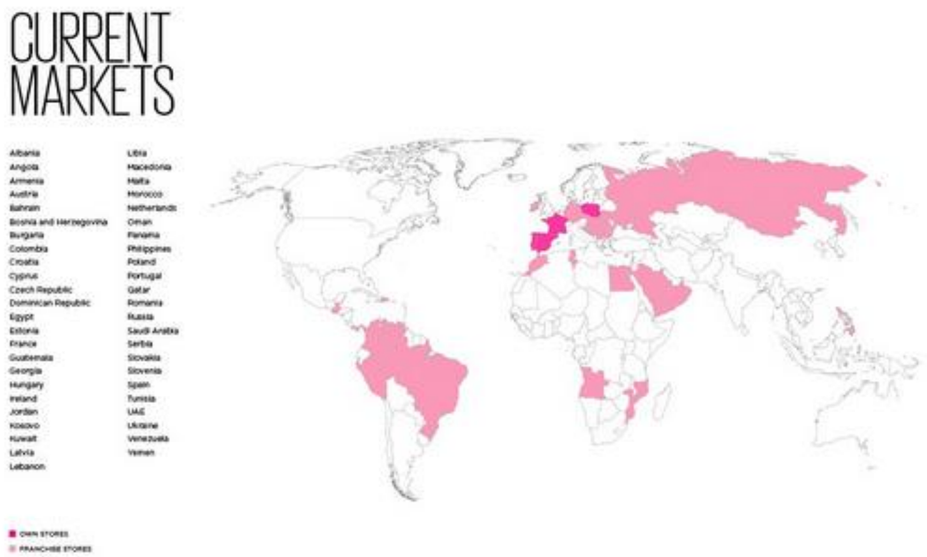
3.1 BUSINESS UNDERSTANDING

This is the phase where the main goal is to analyze and understand what is the business problem, requirements and gain the knowledge to solve the problem with a DM project. Here we must describe the problem, create the plan and define the criteria for a successful accomplishment of the goals (Chapman *et al*., 2000).

In this particular case, the goal is to evaluate consumer's sentiments about Parfois. Parfois is a women's fashion accessories brand created in 1994 by Manuela Medeiros, whose dream was to make exceptional and irresistible fashion accessories, affordable to everyone. The main focus of the company is lying on fashion and following of fashion trends. The brand´s vision is to be the best fashion accessories brand, wherever Parfois decides to operate. With this goal in mind Parfois offers: a huge trendy collection of accessories (designers´ team, based in Oporto and Barcelona create 3.500 SKU´s – Stock Keeping Units – per season; every week the client finds novelties in the store); good relation quality/price; clear visual merchandising; excellent customer service (Metelo, 2015; Parfois Facebook, 2017).

At this moment Parfois has more than 800 stores in more than 75 countries.

Figure 7: List of countries where Parfois have stores



The average annual growth of Parfois is 29%. In some inquires made to Parfois customers we have some values like: 70% visit the store at least twice a month; or 36% visit the store at least once a week. These numbers reveal a high loyalty by Parfois customers.

Parfois, as well as most brands, is present in SMP. In Parfois, the posts in Facebook page have the purpose to announce new campaigns, announce new collections and to advertise Parfois products. Parfois goal when using SMP is to make followers/fans aware about new campaigns, new products and therefore increase the sales and with ewom introduce the brand to new customers and advertise new products to the brand's costumers according the Social Media responsible.

This research will focus in the online platforms, more in particular in Facebook as the page aggregates fans / followers from all over the world. All these fans / followers or potential customers can preview new products, campaigns and comment in the SMP what they think about the products and what they feel about the brand. The SA allows to understand global sentiments about the brand, and sentiments present in each comment in each post. With this analysis, we can better understand what present and future customers think about Parfois and Parfois products.

As we already mentioned, Parfois has Franchise and Consignee stores. As so, all of them have Parfois pages in Facebook too, even some stores have their own Facebook page. It was choosen to use Parfois Facebook page. As it exists one Parfois page for country, it was used the page managed by Parfois marketing department, as it has comments from fans / followers from all over the world.

Our purpose is to extract and analyze posts and comments from Parfois Facebook page and answer to the questions raised by the objectives. Another objective is to create a classification model that could predict the sentiment presented in Parfois Facebook posts with an accuracy of 60%.

3.2 DATA UNDERSTANDING

This stage starts with the extraction, description and verification of quality of the data (Chapman *et al.*, 2000). For this research, it was extracted a set of data with a timeline comprehended between 31th December of 2010 and 1st of March of 2017. This set contained 1.845 posts, 8.256 comments and more than 500.000 reactions from Parfois Facebook page. The data collected is public, so it is available to any private user or brand employee, from Parfois or competitor brand.

3.2.1 DATA EXTRACTION PROCEDURE

The extraction of the data is possible as the SMP used, Facebook, gives access to their Application Programming Interface (API) and with the access to API it is possible to extract, collect and use the data to their own purposes. For the research purposes, it was decided to implement with open source tools an extraction tool.

The first step for this process involved a research how to extract data from Facebook. We found Facebook Graph API. This API it is a low-level HTTP API that allows us to perform various tasks, being one of them the extraction of data from Facebook platform. In general terms, this API provides all core functionalities of Facebook application using Rest web services. For the goal of this research, it was used the Feed connection of the Page object. The Feed represent the list of all Post objects, which contains details like post date, likes, comments, reactions, products advertised, and shares.

Still, this API has some issues, as for example the total number of calls made to the API. When we make consecutive requests to the API, the API automatically blocks our access to make more requests. For example, if we want to get the total number of posts of a page, it will be blocked at the same time because the number of requests will exceed the maximum request limit imposed by Facebook API. So, the workaround its extract the data with time intervals, as the number of requests will reduce drastically.

Along with the API we used also a Java client called RestFB. This API is a very minimal and simply way to fetch information from Facebook and even publish into the application new items. With this client, the access to the core functions of the Facebook API is simplified and even we can use it, to automatize the extraction of the data. For the automation, we started a Java project where we developed an application using RestFB API, that in turn uses Facebook Graph API. Using RestFB, anyone can make a custom implementation, as we made one in order to adapt for our research purpose.

In order to know what we did, we will insert here the main lines of code that we used to extract data from Facebook.

In this line, is the API Call using the version 2.6 of Facebook API:

FacebookClient facebookClient= new

DefaultFacebookClient(Constants.MY_ACCESS_TOKEN, Version.VERSION_2_6);

Calling the feed of the page we want, in this case, *Parfois* Facebook page (official page of *Parfois*) with date range parameters. Due the limits of number of calls of the Facebook API we separated the data collection by years. So, for example, if we want all posts from year 2016 we manage that configurations like this:

```
SimpleDateFormat dateFormatRange = new SimpleDateFormat("dd/MM/yyyy");

Date initDate = dateFormatRange .parse("01/01/2016");

Date endDate = dateFormatRange.parse("01/01/2017");
```

So, the time interval defined is one year.

```
Connection<Post> feed = facebookClient.fetchConnection("Parfois/feed", Post.class,
                                  Parameter.with("since", initDate),
Parameter.with("until", initDate),Parameter.with("limit", 1));
```

After the post's fetch, we start iterating the Feed in order to collect all posts from the range of date we parametrized before:

```
for (List<Post> feedConnectionPage : feed) {
        for (Post post : feedConnectionPage) {
```

Number of shares of the post

```
numberShares = getShares(post.getId(),facebookClient);
```

Product advertised in the post

```
product = searchProduct(post.getMessage());
```

Date of creation of the post

```
postDate = getCorrectDate(post.getCreatedTime());
```

In the next line of code, we are fetching the reactions of the post:

```
reactionsConn = facebookClient.fetchConnection(post.getId()+"/reactions",
                                  Reactions.ReactionItem.class);
```

It was created a function that returns a list with the emotions of the post and as input it receive the fetch result:

```
emotionsList = getEmotions(reactionsConn);
```

Detailing the function, we can see that we are iterating twice, so we can get all information about the reactions of the post:

```
for (List<ReactionItem> reactionsPost : reactionsConn) {
    for (ReactionItem reactionItem : reactionsPost) {
        reaction = new PRFReactions();
        reaction.setReactionId(reactionItem.getId());
        reaction.setName(reactionItem.getName());
        reaction.setType(reactionItem.getType());
        reactionsList.add(reaction);
        countReactions(reactionItem.getType());
        }
}
```

Now it is the time to fetch the comments made in the post. Like before, first we fetch the comments of the post and we call a function to retrieve all comments' information and store it:

```
Connection<Comment> comments =
    facebookClient.fetchConnection(post.getId()+"/comments", Comment.class,
    Parameter.with("fields","from,message,id"));
commentsList = getComments(comments);
```

And now time to store all the post information collected:

```
numberPosts++;
postInformation = new PostInformation(post.getId(), post.getMessage(),numberShares,
numLikes, numWow, numLove, numHaha, numAngry, numSad, commentsList, emotionsList,
product, postDate, onstants.PAGE);
facebookPosts.put(numberPosts, postInformation);
}

}
```

All this information is stored in a structure defined by us, in this particular case:

HashMap<Integer,PostInformation> facebookPosts

We took the option to save all the data into the relational database so we can access information via query or extract it to a file. We made a function where the input is the structure we defined and then we insert into the database the information collected:

facebookPosts.forEach((k,v) -> setDataToInsert(v));

facebookPosts.forEach((k,v) -> setCommentsToInsert(v));

facebookPosts.forEach((k,v) -> setReactionsToInsert(v));

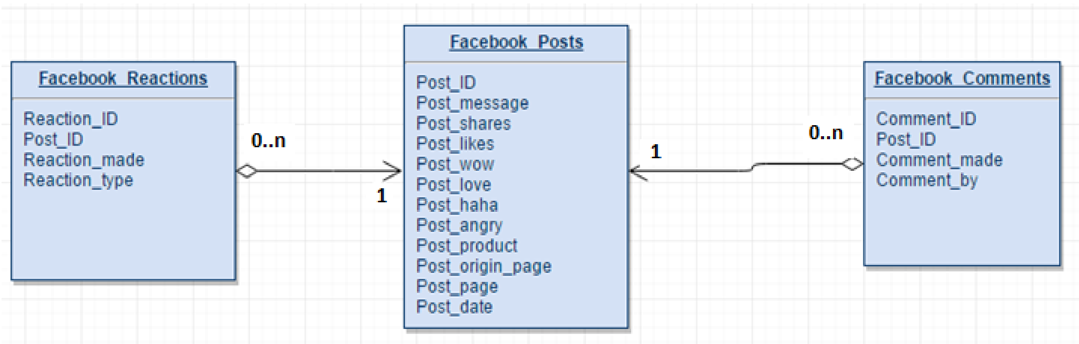
As it could be seen, the posts are being made with a structure, so we had to "personalize" the code of the extraction.

The data extracted is being stored and it is being analysed if the actual structure of the database is correct and if is adequate to our purposes.

All of the data extracted is stored in a relational database (Figure 7) in the next three tables:

- Facebook_Posts: Contains the general data about the post.
- Facebook_Comments: Contains all data about the comments.
- Facebook_Reactions: Includes data collected about the reactions.

Figure 8: Relational database



The general data we store in the relational database is:

i) Post identification (ID): The unique identifier of the post;

ii) Post message: The message that was in the post;

iii) Number of shares: Number of times the post was shared;

iv) Number of total reactions types: Total number of reactions, and total number of reactions type;

v) Post product: Product announced in the post;

vi) Post page: Page where the post was made;

vii) Post date: Date the post was made;

viii) Person: Name of the person who made the reaction or the comment;

ix) Comments: Comments made in the post.

When we iterate each post, we automatically generate two more sets of data, one with the comments and the other with the reactions:

emotionsList = getEmotions(reactionsConn);

commentsList = getComments(comments);

For this research, the comments are the most important set of data, and the one that is used for SA. Maybe it could be questioned why create a tool to extract the data from Facebook, when we have already so many tools that do that. We created the tool using the already enumerated frameworks, because each brand can make the posts as they wish. One of our goals was personalizing the tool, so we can get the maximum information we can. In each post, Parfois puts a link for the website. But it uses the Url shortened by Bitly. So, one of the personalization is getting the long Url, because there we have the id of the product. With this id, we can get all the information about each product announced in each Facebook post. So, with our tool we extracted posts, comments and reactions.

3.2.2 ATTRIBUTES COLLECTED

A Facebook page has different types of information. We have posts and the posts can have comments and reactions made by followers / fans. In order to have a higher volume of data, it was decided to extract data from posts, comments and reactions.

A typical post made by Parfois has a comment, a short link to online shop where we can find the product reference as we can see in figure 8. These posts can be photos or videos.

Figure 9: Parfois Post from Facebook

Posts may have reactions and comments and they are relevant for SA. Therefore, they were extracted too so it could be possible to perform text analysis techniques in the comments.

Figure 10: Comments made in Parfois post

The data collected in Parfois Facebook Page has been classified in three groups: Post attributes, comments attributes and reactions attributes. Table 21 shows the post attributes collected and the next CRISP-DM phase it will be used.

Table 21: Post attributes collected

Attribute	Description	Next CRISP-DM phase
Id	Post identification	Not used
Messages	Message in post	Data preparation
Shares	Total number of shares	Modeling
Likes	Total number of likes	Modeling
Wow	Total number of wow's	Modeling
Love	Total number of love's	Modeling
Haha	Total number of haha's	Modeling
Angry	Total number of angry's	Modeling
Sad	Total number of sad's	Modeling
Product	Products advertised in the post	Data preparation
Origin page	The page where the post was made	Not used
Date	Date of the post	Data preparation
Total Number of comments	Total Number of comments	Modeling
Total number of reactions	Total number of reactions	Modeling

About the reaction attributes collected (Table 22), none of them are used at this research. These attributes are extracted and stored if in the future, the brand wants to make some other metrics, like for example know how many times a person reacted in the page and which products had more attention by a certain user.

Table 22: Reactions attributes collected

Attribute	Description	Next CRISP-DM phase
Reaction Id	Id of the reaction	Not used
Post Id	Id of the post	Not used
Reaction made	Who made the reaction	Not used
Reaction type	Type of reaction made	Not used

Table 23 has the comments attributes collected in the posts of Parfois Facebook page. Extracting these attributes, it is possible to know who, where and what a follower or fan commented in the brand page.

Table 23: Comments attributes collected

Attribute	Description	Next CRISP-DM phase
Comment Id	Id of the comment	Not used
Post Id	Id of the post	Not used
Comment Made	Message in the comment	Data preparation
Comment By	Who made the comment	Not used
Comment Lang	Language of the comment	Not used
Comment Emoji	Emoji in the comment	Data preparation

3.3 DATA PREPARATION

Data preparation phase consists in the data selection, cleaning and formatting for data modeling (Chapman *et al*., 2000). Moreover, in the previous phase we identified attributes requiring some kind of transformation. The output of this phase is the datasets with all the attributes that are used in the modeling phase.

In this phase, it was transformed the data that was already identified as having issues. One of the issues would be cleaning the emoticons, process described in sub-topic 3.3.2. Separate the real comments from the ones which are only mentions of friends in comments is other issue found.

For future work, we are investigating if using Stanford Named Entity Recognizer (NER) will help us to achieve our goal that is discarding or just put apart the comments that are mentions to friends. In the Post attributes (Table 21), we had to prepare for modeling some of them:

- Messages: It was decoded all the message posted by Parfois as we had to identify the purpose of the post, and if it was advertised one or more products we had to separate them in order to identify them later;
- Product: After the message posted by Parfois advertising some products, all of the products id is separated into this field;
- Date: The date is not used in the actual format. We separated the year, month, day and weekday.

Not all the attributes collected from the post's comments are used in the research. The ones used were prepared for modeling:

- Comment made: Comment made in its original language. This attribute is used in data preparation phase, as if is not in English, it will be translated;

- Comment emoji: If the comment contains one or more emoji it is collected into this field;
- Comment EN: Translation of all comments into English.

3.3.1 Emoticons and emojis

The dataset has some issues. Among them is the presence of the following figures in the comments' messages:

😍, 😫, :)))))))), 😟, 🐭, ❤💔💖💄, 😄, 💓💓💓💖💖, 💖, ☺, 👌, 😗, 😂, 💁, 😮, 🙈, ✋💋, 👐💖, 💗, ❤, 👆.

About these kinds of figures, known as emoticons, we must have two concerns: we can clean them from the comments, or we can use them for sentiment analysis (Agarwal *et al*., 2011). Emoticons are facial expressions pictorially represented using punctuation and letters; they express the user's mood. In the present, we have emojis too. Emojis are used much like emoticons and exist in various genres, including facial expressions, common objects, places and types of weather, and animals (Wikipedia, 2016). Table 24 shows some examples of emoticons and their polarity (Pletikosa, Michahelles, 2011)

Table 24: Part of the dictionary of emoticons and emojis

Emoticon	***Polarity***
:-) :) :o) :] :3 :c)	Positive
😍, 💖, ❤	Extremely-Positive
:D C:	Extremely-Positive
:-(:(:c :[	Negative
D8 D; D= DX v.v	Extremely-Negative
😟, 💔	Extremely-Negative
:|	Neutral

Adapted: Pletikosa, Michahelles 2011

Few authors refer to emoticons in their studies. For example, one study concluded that the simplest way to detect the polarity (i.e., positive and negative affect) of a message is based on the emoticons it contains (Hogenboom *et al*., 2013). Emoticons have become popular in recent years, and are now included in all write messages in the social networks. Emoticons are primarily face-based and represent happy or sad feelings, although a wide range of non-facial variations exist: for instance, "<3" represents a heart and expresses love or affection, ":(" is a sad face, ":)" is a happy face (Hogenboom *et al*., 2013).

For our purposes, we can filter the emojis or substitute them with the correspondent emotion. We need to decide which one is the best solution to accomplish our goals. There are some API's we can use to run these tasks and that we can insert in our own API. For gain usage of comments made/having emoticons we must apply other parameters in text tools, when mining comments. Even for irony detection. Is called a Noisy-text Analysis (Saothawan, 2014).

3.3.2 COMMENTS TRANSLATION

Another issue we had were the different languages used in the comments. We had Spanish, Italian, Portuguese, French, English and Arabic.

Following our path in the open source, we used Yandex in order to detect the language used in the comment, as we show in the next lines.

Yandex is a multinational technology company from Russia, specialized in Internet-related services and products. Yandex operates the largest search engine in Russia with about 65% market share in that country and among its services, it has an online translator (Yandex.Translate). This translation services was the one used in this research.

```
if(commentItem.getMessage().equals("")){
  detectedLang = "N/A";
}else{
  detectedLang = Detect.detectLang(commentItem.getMessage());
  }
comment.setLanguage(detectedLang);

if(!detectedLang.equals("en") && !detectedLang.equals("N/A")){
  translatedComm                                                =
Translate.translateComm(commentItem.getMessage(),detectedLang);
}
if (translatedComm!=null && !translatedComm.equals("")){
```

```
comment.setTranslatedMessage(translatedComm);
}else{
  comment.setTranslatedMessage(commentItem.getMessage());
```

So, we first detect the original language of the comment, and then we translate it to English.

```
Translate.setKey(ApiKeys.YANDEX_API_KEY);
Language.getByCode(originLang);
translation = Translate.translateToEN(comment, originLang, Language.ENGLISH);
        }
```

After the translation process made, we have translated almost 90% of the comments. Thise which could not be translated were cases like:

- Linda – confused with the name;
- Amoooooo – Error in words writing;
- My feavory company - Parfois <3;
- Que guay lo quierooooooo.

The next step was translating the 10% that was not translated. Our goal was to have the most translated comments as possible. Therefore, it was extracted to a file all comments and reviewed the ones who were not automatically translated. After that, they were translated manually and added to the file along with the comments already translated automatically.

3.3.3 TEXT ANALYSIS TO RETRIEVE SENTIMENTS AND TYPES

After the translation, we started to make text analysis to retrieve the sentiments and the types from the comments. For types, it is understood the whole set of words that are related to a content, as shown in Table 25.

Although, with the creation of an own dictionary of words for Parfois, it was possible to have the most accurate possible text analysis and thus better achieve the research goals. Table 26 shows the dictionary created with this purpose, as the standards dictionaries of the tool used were not capable of givingthe accuracy needed to achieve the research goals for Parfois, as its business is a specific business.

It was used IBM SPSS Modeler Premium (version 17) as the tool has incorporated text and sentiment analysis tools and predefined dictionaries.

Table 25: Dictionary of types of comments' content

Types	Words
Accessories	hats, glasses, umbrella, watches, belt, bracelet, choker, earrings, loafers, necklace, ring, slippers, suitcase, sun
Parfois	products, brand, online, Parfois, shop, site, store, website
Features	gold, red, colors, khaki, silk, aqua, black, blue, brown, camel, color, colour, cream, fuchsia, gray, green, lime, maroon, olive, purple, silver, teal, white, yellow
Editions	party, Habana, valentine, Christmas, beach, night, fall, spring, summer, winter
Clothing	poncho, scarf, shirt
Wallets	holder, wallet, backpack, bag, laptop bag
Shoes	sandals, shoes, boots
Animals	animal, aardvark, albatross, alligator, alpaca, ant, anteater, antelope, ape, armadillo, baboon, badger, barracuda, bat, bear, beaver, bee, bison, boar, buffalo, butterfly, caribou, cat, caterpillar, cattle, chamois, cheetah, chicken, chimpanzee, chinchilla, chough, clam, cobra, cockroach, cod, cormorant, coyote, crab, crane, crocodile, crow, curlew, deer, dinosaur, dog, dogfish, dolphin, donkey, dotterel, dove, dragonfly, duck, dugong, dunlin, eagle, echidna, eel, eland, elephant, elephant seal, elk, emu, falcon, ferret, finch, fish, flamingo, fly, fox, frog, fur, galago, gaur, gazelle, gerbil, giant panda, giraffe, gnat, gnu, goat, goldfinch, goldfish, goose, gorilla, goshawk, grasshopper, grouse, guanaco, guinea fowl, guinea pig, gull, hamster, hare, hawk, hedgehog, heron, herring, hippopotamus, hornet, horse, hummingbird, hyena, jackal, jaguar, jay, jellyfish, kangaroo, koala, komodo dragon, kouprey, kudu, lapwing, lark, lemur, leopard, lion, llama, lobster, locust, loris, louse, lyrebird, magpie, mallard, manatee, marten, meerkat, mink, mole, monkey, moose, mosquito, mouse, mule, narwhal, newt, nightingale, octopus, okapi, opossum, oryx, ostrich, otter, owl, ox, oyster, panther, parrot, partridge, peafowl, pelican, penguin, pheasant, pig, pigeon, pony, porcupine, porpoise, prairie dog, quail, quelea, rabbit, raccoon, rail, ram, rat, raven, red deer, red panda, reindeer, rhinoceros, rook, ruff, salamander, salmon, sand dollar, sandpiper, sardine, scorpion, sea lion, sea urchin, seahorse, seal, shark, sheep, shrew, shrimp, skunk, snail, snake, spider, squid, squirrel, starling, stingray, stinkbug, stork, swallow, swan, tapir, tarsier, termite, tiger, toad, trout, turkey, turtle, vicuña, viper, vulture, wallaby, walrus, wasp, water buffalo, weasel, whale, wolf, wolverine, wombat, woodcock, woodpecker, worm, wren, yak, zebra

For sentiments analysis, we used the dictionary provided by the software which classifies the comments sentiments in negative and positive. Within the positives and negatives it were incorporated sub-categories. All this structure that classifies the comments is represented in Table 26.

Table 26: Sentiment categories used in IBM SPSS Modeler

Sentiment	Category	Sub-category
Negative		
	Negative/General Negative	
		<NegativeFeeling>
		<NegativeBudget>
		<Negative>
		<NegativeFunctioning>
		<NegativeRecommendation>
		<NegativeCompetence>
	Negative/NegativeFeeling	
		<NegativeFeeling_Emoticon>
		<NegativeFeeling>
	Negative/NegativeAttitude	
		<NegativeAttitude>
Positive		
	Positive/General positive	<PositiveBudget>
		<Positive>
		<PositiveCompetence>
		<PositiveRecommendation>
		<PositiveFunctioning>
	Positive/PositiveFeeling	
		<PositiveFeeling_Emoticon>
		<PositiveFeeling>
	Positive/PositiveAttitude	
		<PositiveAttitude>

In addition, it was set up in IBM SPSS Modeler categories related with Parfois core business as Customer Service, categories of products, budget, features of the products, the context of the products and categories associated to Parfois itself. With these categories, it would be possible to identify the words in text mining and associate them to the correspondent categories, as show in Table 27.

Table 27: Categories and sub-categories used

Category	Sub-category
Customer_Service	
	<CustomerSupport>
	<Website>
	<Follow-Up>
	<email>
<Parfois>	
	<Location>
	<Parfois>
	<Store>
	<url>
<Budget>	
	<Currency>
	<Budget>
<Animals>	
<Person>	
<Shoes>	
<Accessories>	
<Contextual>	
<Hashtag>	
<Editions>	
<Clothing>	
<Features>	
<Wallets>	

3.3.4 OTHER TRANSFORMATIONS

As the data extracted is too unstructured, it was necessary to prepare the data extracted, so it could be possible to have the most amount of data, so the SA could be the most accurate and have the most complementary information. All this information can help to have the most detailed results as possible.

Table 28: Variable transformation

Variable	Meaning	Transformation
Post Type	Type of post.	Verified the post type. Video, Photo, Link.
Post Reason	Reason why the post was made	Verified manually the reason to make the post. Sales, Job offer, New opening, etc.
Season	Season of the product	Made a Vlookup with the excel containing all products of the online platform
Product category	Category of the product	Made a Vlookup with the excel containing all products of the online platform
Year	Year of the post	Separated the year of the post from the date
Month	Month of the post	Separated the month of the post from the date
Day	Day of the post	Separated the day of the post from the date
Weekday	Day of the week where the post was made	Calculated the weekday when the post was made
Product	Identify each product of the post	Made a Vlookup with the excel containing all products of the online platform
Number of products	Number of products in post	Sum of all products presented in the post
Number products dif	Number of different products in post	Separate the different number of products in the post
Products diff	Exists different products	Calculated based in the previous variable
Days_last_post	Days between posts	Calculates the days remaining between each post
N_Posts_Day	Number of posts per day	Counts all posts made in the same day
More_Than_one_post_day	More than a post in a day	Flag to verify it was made more than one post in that day
Public Agreement	Post public agreement	$\frac{Likes + Wow + Love + Haha}{Average(Reaction\ number)}$
Popularity	Popularity of the post content	$\frac{Post\ number\ of\ shares}{Average(shares\ number)}$
Public Involvement	Post public involvement	$\frac{Post\ number\ of\ comments}{Average(comments\ number)}$
Positive weight	Weight of the positive sentiments in the post	$\frac{Positive\ comments\ number}{Total\ number\ of\ comments)}$
Negative weight	Weight of the negative sentiments in the post	$\frac{Negative\ comments\ number}{Total\ number\ of\ comments)}$
Global weight	Global sentiment weight of the post	Positive weight - Negative weight

3.3.5 Attributes for Modeling

The following tables show all the attributes used for modeling process. Some attributes were gathered directly in the extraction, and other were calculated from the data extracted. It was collected the products attributes and calculated the attributes for modeling phase as listed in Table 31. For products attributes we have the season or the category of the product associated to the post.

Table 29: List of products attributes for modeling

Attribute	Description	Type of attribute	Objectives
Products attributes			
Season	Season of the product	Calculated Nominal	Calculate the percentage of posts with a product of a season
Belts	Product in the post	Calculated Flag	Calculate the percentage of posts with belts
Watches	Product in the post	Calculated Flag	Calculate the percentage of posts with watches
Wallets	Product in the post	Calculated Nominal	Calculate the percentage of posts with wallets
Shoes	Product in the post	Calculated Nominal	Calculate the percentage of posts with shoes
Bijou	Product in the post	Calculated Nominal	Calculate the percentage of posts with bijou
Textiles	Product in the post	Calculated Nominal	Calculate the percentage of posts with textiles
Bags	Product in the post	Calculated Nominal	Calculate the percentage of posts with bags
Night Bags	Product in the post	Calculated Nominal	Calculate the percentage of posts with night bags
Party Bags	Product in the post	Calculated Nominal	Calculate the percentage of posts with party bags
Sunglasses	Product in the post	Calculated Nominal	Calculate the percentage of posts with sunglasses
Travel	Product in the post	Calculated Nominal	Calculate the percentage of posts with travel
Hair Art.	Product in the post	Calculated Nominal	Calculate the percentage of posts with hair articles
Clothes	Product in the post	Calculated Nominal	Calculate the percentage of posts with bags
Hats	Product in the post	Calculated Nominal	Calculate the percentage of posts with hats
Winter Art.	Product in the post	Calculated Flag	Calculate the percentage of posts with winter articles

Table 30 have the post attributes collected / calculated for modeling as the post type, reason of the post, shares or product (s) associated.

Table 30: List of post attributes for modeling

Attribute	Description	Type of attribute	Objectives
Post attributes			
Post Type	Type of the post made	Gathered Nominal	Calculate the percentage of different post types
Post Reason	Reason of the post	Gathered Nominal	Calculate the percentage of different post reasons
Shares	Number of shares of the post	Gathered Quantitative	Get the number of shares of the post
Product Associated	The post has or not a product associated	Gathered Flag	Calculate the percentage of posts with products associated
Different Products Associated	The post has or not a product associated	Gathered Flag	Calculate the percentage of posts with different products associated
More than One Post during the Day	If the page has more than one post made on the day	Calculated Flag	Calculate the percentage of posts made in the same day
Post_Tot_Comments	Total comments of the post	Gathered Quantitative	Used to calculate the average number of comments
Post_Tot_Reactions	Total reactions made in the post	Gathered Quantitative	Used to calculate the average number of reactions
Number of Products Associated	Number of products in the post	Calculated Quantitative	Calculate the percentage of posts with same number of products associated
Number of Different Products Associated	Number of different products that the post has	Calculated Quantitative	Calculate the percentage of posts with different number of products associated
Number of Days After Last Post	Number of days after the last post in the page	Calculated Quantitative	Calculate the number of days after last post
Number of Posts during the Day	Number of the posts made in the page in one day	Calculated Quantitative	Calculate the percentage of posts that were made in the same day

Regarding the attributes of the post associated to dates, Table 31 has the post date attributes collected / calculated for modeling. It was collected the year, month and day. After it was calculated each weekday of the post.

Table 31: List of date post attributes for modeling

Attribute	Description	Type of attribute	Objectives
Date of the post			
Year	Year of the post	Gathered Nominal	Calculate the percentage of posts made in the year
Month	Month of the post	Gathered Nominal	Calculate the percentage of posts made in the month
Day	Day of the post	Gathered Nominal	Calculate the percentage of posts made in the day
Weekday	Weekday of the post	Calculated Nominal	Calculate the percentage of posts made in the weekday

The reactions allowed in Facebook and made by fans / followers in the Parfois posts, that were collected and calculated for modeling are listed in Table 32.

Table 32: List of fans / follower's reactions for modeling

Attribute	Description	Type of attribute	Objectives
Reactions by Followers			
Likes	Number of Likes of the post	Gathered Quantitative	Used for positive reactions calculation
Wow	Number of wow's of the post	Gathered Quantitative	Used for positive reactions calculation
Love	Number of love's of the post	Gathered Quantitative	Used for positive reactions calculation
Haha	Number of Haha's of the post	Gathered Quantitative	Used for positive reactions calculation
Angry	Number of Angry's of the post	Gathered Quantitative	Used for negative reactions calculation
Sad	Number of Sad's of the post	Gathered Quantitative	Used for negative reactions calculation

It was calculated a set of attributes regarding the reactions of posts and sentiments associated to posts. All these values were necessary to understand and calculate the sentiments associated to posts and are important to calculate the adequate metrics for decision makers. These attributes are show in Table 33.

Table 33: List of reactions for modeling

Attribute	Description	Type of attribute	Objectives
Reactions			
Negative Reactions	Number of negative reactions	Calculated Quantitative	Total of negative reactions
Positive Reactions	Number of positive reactions	Calculated Quantitative	Total of positive reactions
Public Agreement	Number of likes of a specific post in relation to average number of likes	Calculated Quantitative	Calculates how much a post is publicly accepted.
Popularity	Number of shares of a specific post in relation to average number of shares	Calculated Quantitative	Calculates the popularity of the post
Public Involvement	Number of comments of a specific post in relation to average number of comments	Calculated Quantitative	Calculates how much a user show publicly their opinion
Post_Sentiment	Flag if the post has a sentiment	Calculated	Calculates if exist sentiment in the post
Negative	If the post has a negative sentiment	Calculated	Used to calculate the percentage of negative posts
Positive	If the post has a positive sentiment	Calculated	Used to calculate the percentage of positive posts
Positive reactions	Sum of post positive reactions	Calculated	Calculate the overall positive reactions
Negative reactions	Sum of post negative reactions	Calculated	Calculate the overall negative reactions

3.4 MODELING

Modeling is the selection and parametrization of modeling techniques, such as the cluster analysis and the decision tree algorithms (Chapman et al., 2000).

In order to predict sentiments, we used decision tree algorithms. Ayoubloo *et al*. (2011: p. 10115) explain that a decision tree is nothing more than an algorithm which will test in every step of the analysis a determined condition. If we imagine a tree, each of the tree branches is a choice and each leaf is classification of the choice.

From the various algorithms used to build the decision trees we highlight Classification and Regression Tree (CART) or C5.0 (Delen *et al*., 2013).

For now, a brief description of the decision trees algorithms:

- Cart (Breiman et al., 1984) can predict continuous dependent variables, known as regression analysis and also predict categorical variables, known as classification analysis (Brezigar-Masten & Masten, 2012). The algorithm begins splitting the data in two subsets partitioning both. This process is recursive and is made till is not more possible and always try to homogenize each subset. Breiman *et al*. (1984) recommends using the Gini Index, that is the process to reduce the impurity. The algorithm is simple to understand by whom will use it (Li *et al*., 2010).
- C5.0 (Quinlan, 1993) is an improvement of the C4.5 decision tree algorithm. It can handle any combination with any combination. It handles with numeric and nominal attributes, it is able to train data with missing values and it accepts noisy data. Is also able to learn disjunctive expressions and have a process to solve the problem of the over-fitting (Polat & Gunes 2009; Mantas & Abellán, 2014).

The models obtained with the algorithms can be combined in order to improve the model's quality. For this purpose, can be used the bagging which aims to increase model generality; or boosting which aims to increase the model's accuracy.

If there is no algorithm that is more suitable to classify the feeling of the post in positive or non-positive, can be used the algorithms CART and C5 with different parametrizations (Tables 34 and 35).

Table 34: CART models used

	Model A	**Model B**	**Model C**
	CART	**CART boosting**	**CART bagging (voting)**
Tree depth	5	5	5
Minimum records in parent branch	2	2	2
Minimum records in child branch	1	1	1

Table 35: C5 models used

	Model D C5 favour accuracy	Model E C5 favour generality	Model F C5 boosting favour accuracy	Model G C5 boosting favour generality
Tree depth	5	1	2	1
Prune Severity	75	75	75	75
Minimum records in child node	1	1	1	1

Since trees are not always able to make an accurate classification, in this research we also use artificial neural networks with the algorithm Multilayer Perceptron (MLP). This algorithm uses different Multilayer Feed Forward neural networks. This type of networks can be trained in a different decision surfaces in the input space. MLP training processes have a limitation, that consists in large computation consumption (Ihsan *et al*., 2017).

In this research, it was used the backpropagation algorithm. The algorithm searches the minimum error function in the weight space using the gradient descent method. The solution of the problem is the combination of the weights that minimizes the error function. It is called back propagation because the error is calculated at the output and it is distributed through the network (Sietsma and Dow, 1991).

It was also used three different parameterizations as Table 36 shows:

Table 36: Models parametrizations

	Model H ANN	Model I ANN boosting	Model J C5 boosting favour accuracy
Model type	Multilayer Perceptron (MLP)	Multilayer Perceptron (MLP)	Multilayer Perceptron (MLP)
Hidden layer	1	1	1
Hidden layer neurons	6	-	-
Overfit prevention set	30%	30%	30%

3.5 EVALUATION

In this phase occurs an evaluation to the model and model construction, to make sure that every step is correct and that nothing is missing in the model design. When these steps are totally guaranteed, the project leader must decide what to do with the results that were produced in the research (Chapman *et al*., 2000).

Classification matrix or confusion matrix is the name of a tool that evaluates the performance of the model used in classification methods. The matrix has two entries. These entries are the predicted and the actual classes from the model (Delen *et al*., 2013; IBM, 2016a).

Table 37 represents an example of a classification matrix, where the true positives are classified correctly and the true negatives are the ones classified correctly. The false positives were incorrectly classified and the false negatives are the positives incorrectly classified (IBM, 2016a).

Table 37: Classification matrix

		Predicted sentiment	
		Successful (positive sentiment)	Unsuccessful (non positive sentiment)
Observed sentiment	Successful (positive sentiment)	True positives	False negatives
	Unsuccessful (negative sentiment)	False positives	True negatives

Adapted: Delen et al. (2013).

Regarding this research, predicting the posts which will have positive sentiments was considered the true positives.

It was used in this research the following metrics: Accuracy; Sensitivity; Specificity; and AUC.

Accuracy is data percentage correctly classified by the model, so, it is the ratio of the cases predicted correctly to the total number of cases (Delen *et al*., 2013; IBM, 2016a).

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{True Negatives} + \text{False positives} + \text{False Negatives}}$$

Sensitivity is the function which gives the rate of the true positives correctly identified and classified (He & Garcia, 2009; Delen *et al*., 2013).

$$\text{Sensitivity} = \frac{True\ Positives}{True\ Positives + True\ Negatives}$$

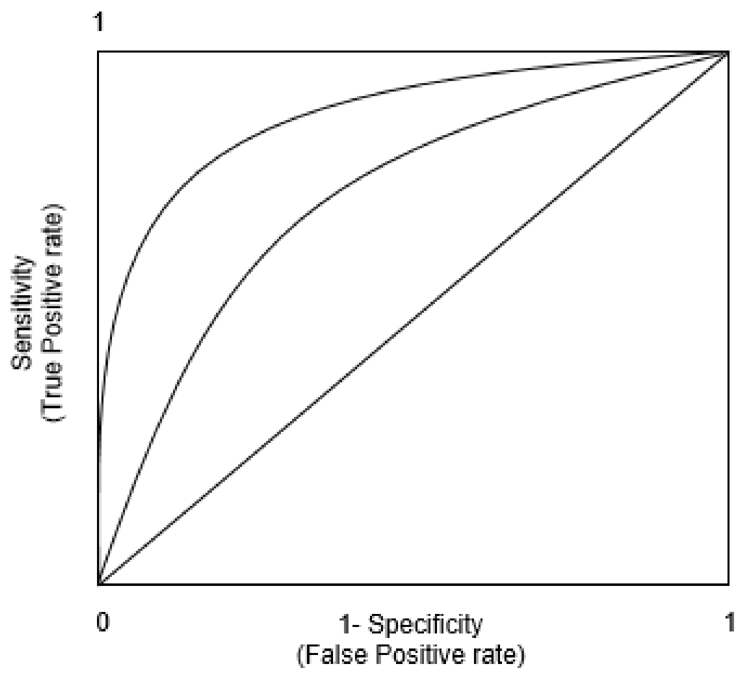
Specificity is the percentage of true negatives well identified and classified (Delen *et al*., 2013).

$$\text{Specificity} = \frac{True\ Negatives}{True\ Negatives + False\ Positives}$$

However, for the accuracy calculation, it was used a ROC (Receiver Operation Characteristic) curve, as shown in Figure 11. It is a plot showing two probabilities, namely sensitivity and 1-specificity (Pundir & Amala, 2015). It is used to visualize the trade-off between the rate at which the model can accurately recognize the positive tuples versus the rate at which it identifies the negative tuples as positive. This process is made in different portions of the test set.

To measure the accuracy of the model it will be used AUC and an area of 1.0 represents the perfect accuracy of the model (Pundir & Amala, 2015; IBM, 2016a).

Figure 11: ROC Curve representation



Source: Sayad (2016).

Additionally, it should be noted that the evaluation of the models should be based on data considered to be clean. Thus, the total sample of posts is randomly divided into training sample (70%) and test sample (30%). In order for the model to be generalized, it is necessary to present good quality in the test sample.

3.6 DEPLOYMENT

The deployment phase of this research will be the full work made in the thesis and the analysis made. It would be necessary to present to internal stakeholders of Parfois this research and discuss with them the best approach to the brand. This research also intends to be an added value to retail business, and help them understand the benefits to analyze the consumers sentiments when commenting in the SMP of the brands.

4 RESULTS AND DISCUSSION

In this chapter is analyzed and evaluated the methods developed in the modeling phase of the CRISP-DM.

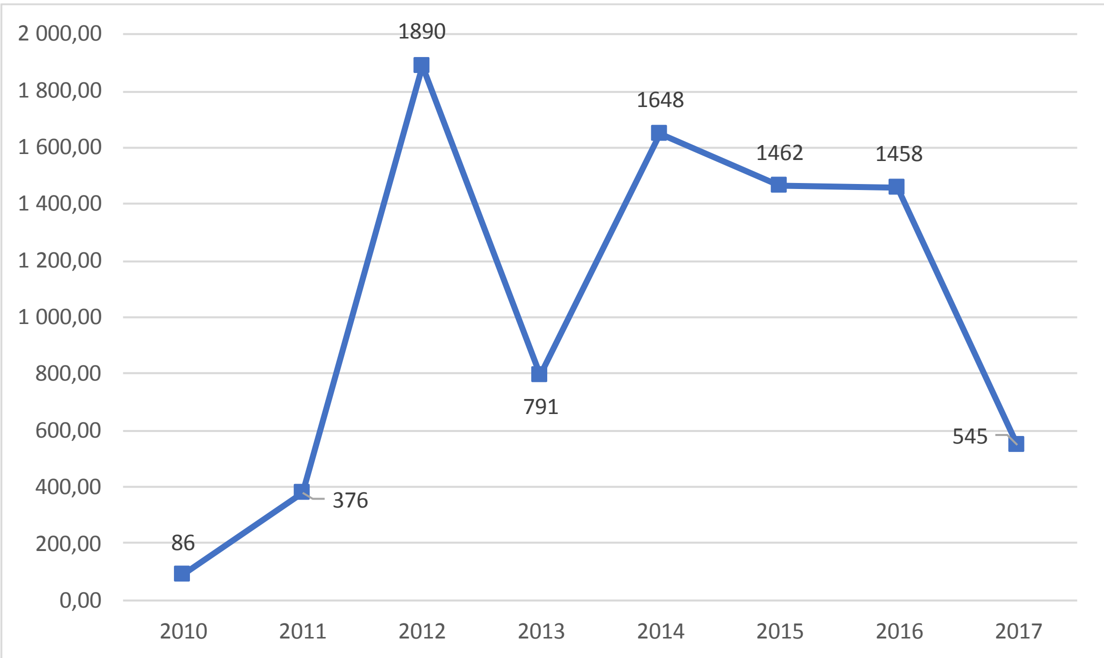
The results presented in this chapter were achieved through all the processes describes in the previous chapters. All the data, structured or unstructured, was extracted and transformed into meaningful information, with the help of dictionaries. Qualitative and quantitative data were combined and processed in order to achieve the defined objectives.

4.1 COMMENTS CHARACTERIZATION

The 8.256 comments extracted from Parfois are interactions from the fans / followers with the brand. Using text mining techniques, it was possible to extract the sentiment and other content that were present in each comment.

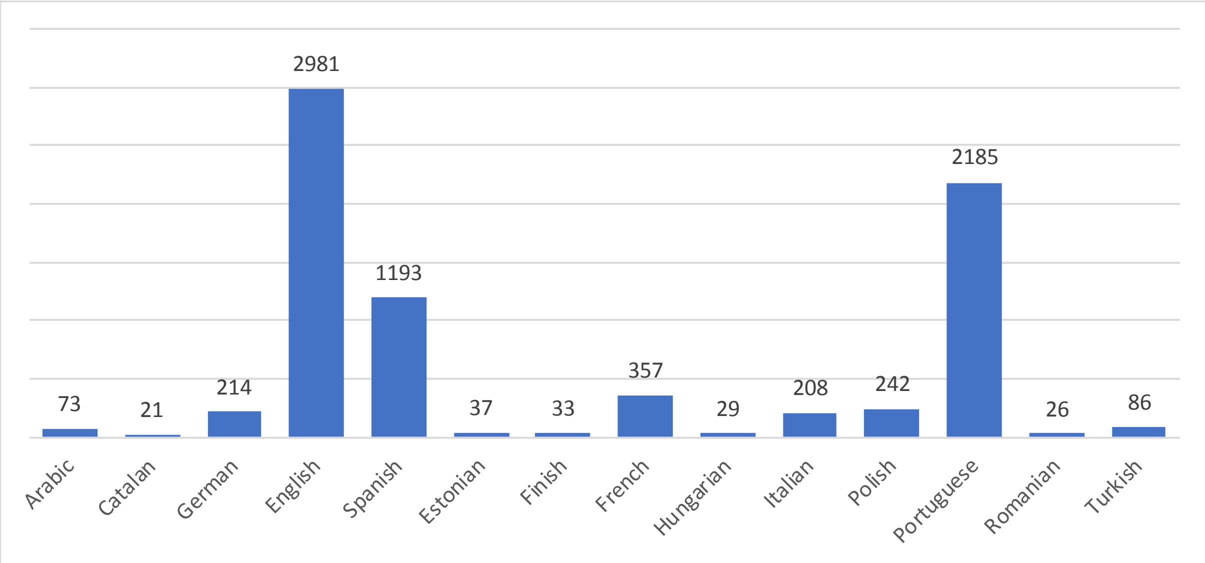
Figure 12 shows that the years with more comments are 2012 (22.9%) and 2014 (22.0%). Although, in 2017 just for the period from January to March the comments were extracted.

Figure 12: Comments sentiments distribution by year



When analyzing the distribution of the comments by the language of the comment (Figure 13), we notice that the higher number of comments are made in English (38.8%), Portuguese (28.4%) and Spanish (15.5%).

Figure 13: Comments distribution by language used



4.1.1 CONTENTS

Figure 14 shows the global word cloud that resulted from the TM process. Words like love, store, Parfois, buy, want or already are more frequent in the comments message.

Figure 14: Global Word Cloud

Regarding types of contents present in the comments message (Table 38), the most frequent is Parfois (15.1%), followed by person (13.2%) and features (11.3%). It is important to note that 35.2% of the comments have no associated content type, that is, comments with links, referring another person or commenting with emojis.

Table 38: Types of contents distribution

Type	Comments	%
Accessories	128	1.55
Animals	129	1.56
Budget	552	6.69
Clothing	25	0.30
Contextual	279	3.38
Editions	32	0.39
Features	934	11.31
Wallets	409	4.95
Hashtag	11	0.13
Parfois	1249	15.13
Person	1093	13.24
Shoes	90	1.09
Customer Service	417	5.05
Without type	2908	35.22
Total	**8256**	**100.00**

4.1.1 SENTIMENTS

Table 39 shows that only 49.2% of the comments have either a positive or a negative sentiment, being the positive (44.88%) almost the half of the total of the comments. Moreover, the comments with a negative sentiment expressed are residual (only 4.32%).

Table 39: Comments sentiment distribution

Sentiment	Comments	%
Negative	357	4.32
Neutral	4194	50.80
Positive	3705	44.88
Total	**8256**	**100.00**

Regarding the distribution of the sentiments by year (Table 40) evidences that the years 2011, 2014 and 2016 are the years with more negative comments. When analyzing the positive comments, the years 2010 and 2013 have the most positive comments, with percentages around 50%.

Table 40: Comments sentiments distribution by year

		Sentiment							
		Negative		Neutral		Positive		Total	
		Comments	%	Comments	%	Comments	%	Comments	%
Year	2010	2	2.33	37	43.02	47	54.65	**86**	**100.00**
	2011	36	9.57	152	40.43	188	50.00	**376**	**100.00**
	2012	38	2.01	908	48.04	944	49.95	**1890**	**100.00**
	2013	24	3.03	348	43.99	419	52.97	**791**	**100.00**
	2014	93	5.64	835	50.67	720	43.69	**1648**	**100.00**
	2015	51	3.49	735	50.27	676	46.24	**1462**	**100.00**
	2016	92	6.31	859	58.92	507	34.77	**1458**	**100.00**
	2017	21	3.85	320	58.72	204	37.43	**545**	**100.00**

4.1.2 RELATIONSHIP BETWEEN CONTENT AND SENTIMENT

Along with the analysis some networks created. These networks connect types of content commented with the sentiments presented in the comments. Figure 15 shows the global network, with the types and sentiment associated to that type. Types of content like Parfois, contextual or features leads to positive sentiments. Instead, type of contents as animals, person or budget leads to negative sentiments.

The Figures 15, 16 and 17 translate networks that present correlations of categories that identifies related topics mined in SMP. The categories are associated with each other by fans / followers and are present in their comments. Thin lines are low correlations and thick lines are strong correlations.

Figure 15: Global network linking contents and sentiments

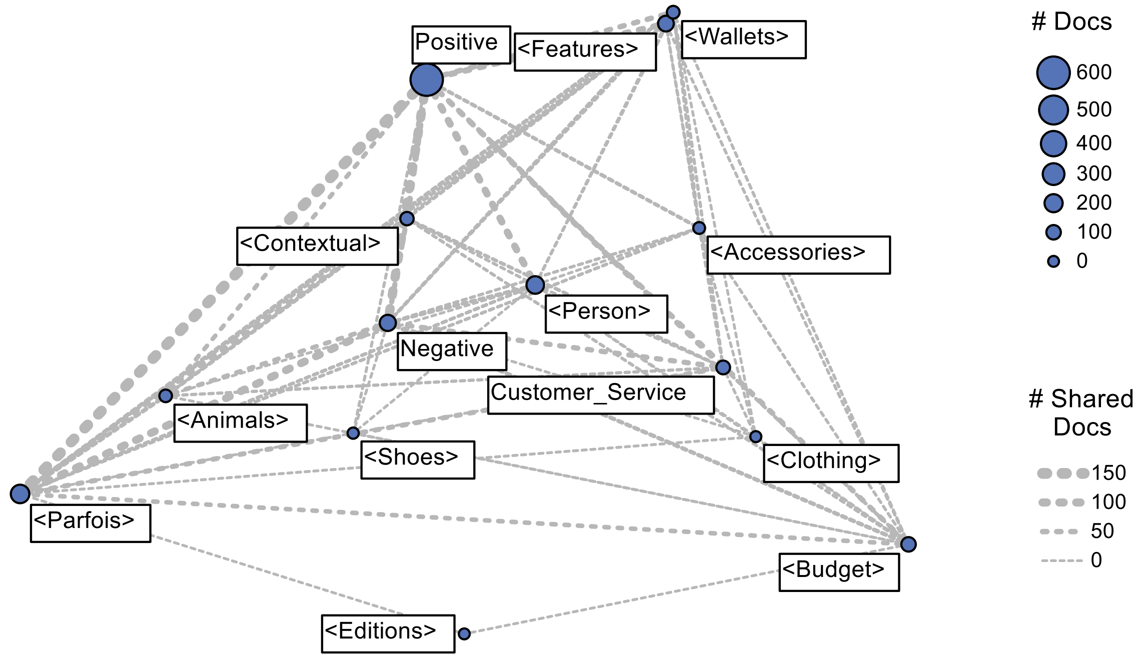


Figure 15 shows the negative network and presents which types of content lead to negative sentiments. Parfois is the type of content more associated to the negative sentiment. Also, wallets, contextual, customer services and animals are connected to the negative sentiment.

Figure 16: Network linking contents and negative sentiments

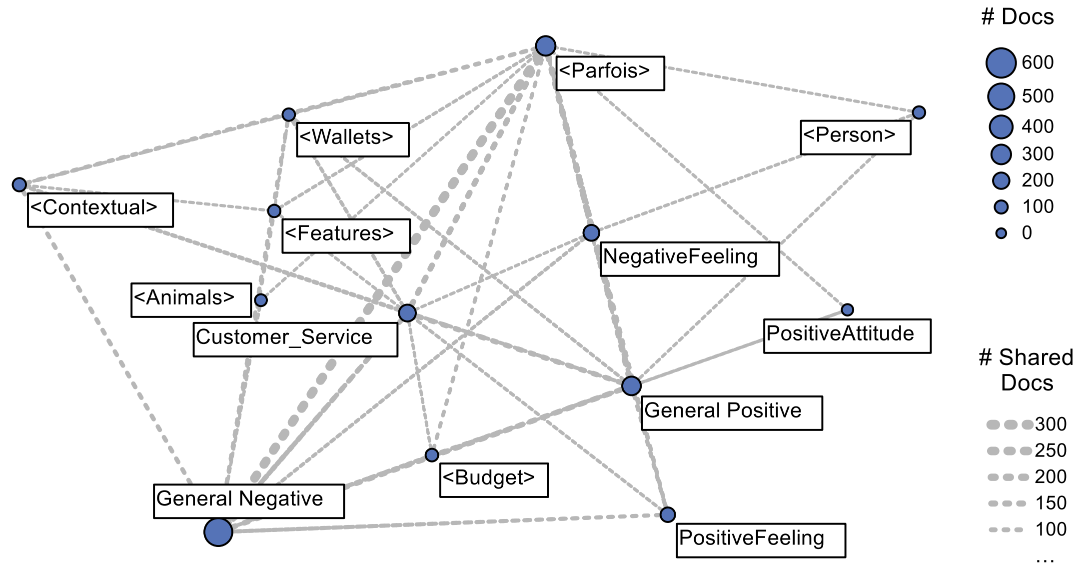


Figure 16 shows the positive network and presents which types lead to positive sentiments. This network focuses the positive sentiments and it can be observed that Parfois and person lead to positive / general positive sentiments. Customer service and contextual lead to positive / positive attitude. Parfois and wallets are associated as it wallets and features. Both of these associations lead to a positive sentiment. Parfois also lead to negative feeling, so the sentiments of fans / followers are distributed in both negative and positive sentiments.

Figure 17: Network linking contents and positive sentiments

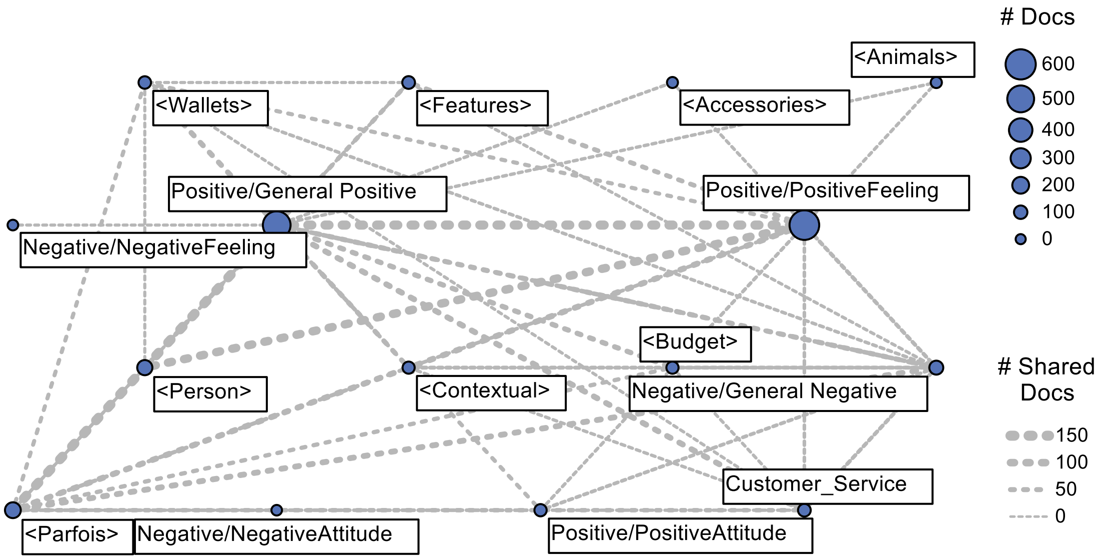


Table 41 shows the distribution of the sentiments by type of content.

Table 41: Sentiment distribution by type of content

Type of content	Negative		Neutral		Positive		Total	
	Comments	%	Comments	%	Comments	%	Comments	%
Accessories								
No	352	4.33	4143	50.97	3633	44.70	8128	100.00
Yes	5	3.91	51	39.84	72	56.25	128	100.00
Animals								
No	312	3.84	4140	50.94	3675	45.22	8127	100.00
Yes	45	34.88	54	41.86	30	23.26	129	100.00
Budget								
No	336	4.36	3797	49.29	3571	46.35	7704	100.00
Yes	21	3.80	397	71.92	134	24.28	552	100.00
Clothing								
No	354	4.30	4185	50.84	3692	44.85	8231	100.00
Yes	3	12.00	9	36.00	13	52.00	25	100.00
Contextual								
No	315	3.95	4061	50.91	3601	45.14	7977	100.00
Yes	42	15.05	133	47.67	104	37.28	279	100.00
Editions								
No	356	4.33	4181	50.84	3687	44.83	8224	100.00
Yes	1	3.13	13	40.63	18	56.25	32	100.00
Features								
No	310	4.23	3638	49.69	3374	46.08	7322	100.00
Yes	47	5.03	556	59.53	331	35.44	934	100.00
Hashtag								
No	357	4.33	4185	50.76	3703	44.91	8245	100.00
Yes	0	0.00	9	81.82	2	18.18	11	100.00
Parfois								
No	244	3.48	3568	50.92	3195	45.60	7007	100.00
Yes	113	9.05	626	50.12	510	40.83	1249	100.00
Person								
No	316	4.41	3476	48.53	3371	47.06	7163	100.00
Yes	41	3.75	718	65.69	334	30.56	1093	100.00
Shoes								
No	353	4.32	4149	50.81	3664	44.87	8166	100.00
Yes	4	4.44	45	50.00	41	45.56	90	100.00
Wallets								
No	323	4.12	4031	51.37	3493	44.51	7847	100.00
Yes	34	8.31	163	39.85	212	51.83	409	100.00
Customer Service								
No	273	3.48	3967	50.61	3599	45.91	7839	100.00
Yes	84	20.14	227	54.44	106	25.42	417	100.00

More than 50% of the comments about accessories, clothing, editions and wallets have a positive sentiment. In the other hand, negative sentiments are more relevant in comments about animals (34.9%) and customer service (20.1%).

When analyzing the distribution of the sentiments by the language of the comment (Table 42), we notice that more than 50% of the comments made in German or Polish have positive sentiments. Romanian, Spanish and Portuguese languages have the higher percentage of comments with negative sentiments (percentages are slightly above 5%).

Table 42: Sentiment distribution by comments' language

Language	Negative		Neutral		Positive		Total	
	Comments	%	Comments	%	Comments	%	Comments	%
Arabic	1	1.37	47	64.38	25	34.25	73	100.00
Catalan	1	4.76	13	61.90	7	33.33	21	100.00
German	2	0.93	97	45.33	115	53.74	214	100.00
English	102	3.42	1593	53.44	1286	43.14	2981	100.00
Spanish	78	6.54	560	46.94	555	46.52	1193	100.00
Estonian	1	2.70	21	56.76	15	40.54	37	100.00
Finish	1	3.03	25	75.76	7	21.21	33	100.00
French	17	4.76	195	54.62	145	40.62	357	100.00
Hungarian	0	0.00	18	62.07	11	37.93	29	100.00
Italian	2	0.96	144	69.23	62	29.81	208	100.00
Polish	12	4.96	95	39.26	135	55.79	242	100.00
Portuguese	123	5.63	1027	47.00	1035	47.37	2185	100.00
Romanian	2	7.69	18	69.23	6	23.08	26	100.00
Turkish	1	1.16	61	70.93	24	27.91	86	100.00

4.2 IDENTIFICATION OF THE CHARACTERISTICS OF THE POSTS THAT INFLUENCE THE ASSOCIATED SENTIMENT OF THE POST

4.2.1 POSTS CHARACTERIZATION

Table 43 shows the number of posts made each year, since 2010 to 2017 (up to the end of March). The year with the most amount of posts was 2016 (21.4%) followed by the year of 2014 (21.3%). The results allow to conclude that with the increase of Facebook importance in the society the number of posts increased too. The months with the most percentage of posts are the last 3 months of the year (October, November and December) and the weekend days have slightly less posts than the workdays.

Table 43: Number of posts by date attributes

Date Attributes		Count	%	Mean	Standard deviation	Minimum	Median	Maximum
Year	2010	103	5.58					
	2011	97	5.26					
	2012	238	12.90					
	2013	175	9.49					
	2014	393	21.30					
	2015	362	19.62					
	2016	395	21.41					
	2017	82	4.44					
	Total	**1845**	**100**					
Month	Jan	141	7.64					
	Feb	143	7.75					
	Mar	132	7.15					
	Apr	118	6.40					
	May	107	5.80					
	Jun	128	6.94					
	Jul	155	8.40					
	Aug	135	7.32					
	Sep	159	8.62					
	Oct	241	13.06					
	Nov	183	9.92					
	Dec	203	11.00					
	Total	**1845**	**100.0**					
Weekday	Mon	295	15.99					
	Tue	290	15.72					
	Wed	295	15.99					
	Thu	272	14.74					
	Fri	290	15.72					
	Sat	211	11.44					
	Sun	192	10.41					
	Total	**1845**	**100.0**					
Number of Posts during the Day		1845		1.85	1.77	1.00	1.00	16.00
Number of Days After Last Post		1845		1.41	7.73	0.00	1.00	271.00
More than One Post during the Day	No	993	53.82					
	Yes	852	46.18					
	Total	**1845**	**100**	1.41	7.73	0.00	1.00	271.00

It is also relevant that on average Parfois posts 1.85 posts each day, but not every day of the year. Indeed, 46.2% of the posts are made in days with more than one post during the day.

Table 44 shows the characterization of the posts made in Parfois Facebook page. The posts with photo have the higher percentage (87.9%) and when checking the post reason, the higher percentage is sales (67.1%). This means that Parfois uses photos to show products in order to increase their sales.

Table 44: Post characterization

Post Attributes		**Posts**	**%**
Post Type	Link	142	7.71
	Photo	1617	87.79
	Video	83	4.51
	Total	**1842**	**100**
Post Reason	Announcement	29	1.60
	Blog	157	8.64
	Campaign	241	13.26
	Contest	25	1.38
	Job Offer	21	1.16
	New Store	108	5.94
	Newspaper	16	0.88
	Sales	1220	67.14
	Total	**1817**	**100**

Bijou and bags are the product categories with more posts (Table 45). The others have almost an equality percentage of posts mentions. From all the posts extracted, 59.2% have a product associated and 52.9% have only one product associated. About the season of the product and having Parfois only two seasons we can verify that the percentage is not much different between both seasons. Although, 37.4% of the posts are generic (without an associated season).

Table 45: Distribution of the products associated to posts

Product category		Posts	%
Belts	No	1802	97.67
	Yes	43	2.33
Watches	No	1804	97.78
	Yes	41	2.22
Wallets	No	1795	97.29
	Yes	50	2.71
Shoes	No	1733	93.93
	Yes	112	6.07
Bijou	No	1525	82.66
	Yes	320	17.34
Textiles	No	1716	93.01
	Yes	129	6.99
Bags	No	1610	87.26
	Yes	235	12.74
Night Bags	No	1697	91.98
	Yes	148	8.02
Party Bags	No	1776	96.26
	Yes	69	3.74
Sunglasses	No	1782	96.59
	Yes	63	3.41
Travel	No	1842	99.84
	Yes	3	0.16
Hair Article	No	1822	98.75
	Yes	23	1.25
Clothes	No	1781	96.53
	Yes	64	3.47
Hats	No	1825	98.92
	Yes	20	1.08
Winter Article	No	1831	99.24
	Yes	14	0.76
Product Associated	No	753	40.81
	Yes	1092	59.19
Different Products Associated	Without Product	753	40.81
	No	976	52.90
	Yes	116	6.29
Season	Fall/Winter	625	33.88
	Spring/Summer	530	28.73
	Generic	690	37.40

Note: **1845 posts.**

4.2.2 Evaluation of both the sentiment and the interaction level associated to the Posts

Public involvement indicates how much post triggers fans / followers to publicly show their opinion. Table 46 shows that the median of public involvement is 0.45 and the mean value is 1.00, meaning that almost always a fan / follower interact with the brand. Moreover, 56.0% of the posts have a positive sentiment associated and only 3.6% have a negative sentiment, meaning that Parfois has a bit more than the half value of total posts with positive sentiments. Still we have 40.3% of neutral sentiments.

Table 46: Evaluation of sentiments and public involvement

Sentiment attributes		Posts	%	Mean	Standard deviation	Minimum	Median	Maximum
Post Sentiment	Negative	67	3.63					
	Neutral	744	40.33					
	Positive	1034	56.04					
	Total	**1845**	**100.0**					
Post Sentiment Present	No	665	36.04					
	Yes	1180	63.96					
	Total	**1845**	**100.0**					
Public Involvement		1845		1.00	1.63	0.00	0.45	20.78

When analyzing the interaction attributes (Table 47) we verify like is by far the most reaction used in the interaction with the posts (on average, posts have 281.2 likes). Negative reactions are residual as it is possible to see when checking the values of the mean (0.02) and median is 0.00. If we use Table 18 metric values public involvement is low, public agreement and popularity are neutral. These metrics were created in another context and another type of brand, not a fashion brand. So, this classification, must be taken in consideration. Like in literature revision, it is advised to adapt the metrics to each company, as their context and brand type are different.

When observing the popularity of the posts from Parfois in Facebook it is visible that the median value is 0.66 and the mean value is 1.00. Regarding public agreement, both values are also low (median=0.67, mean=1,00), meaning that both public agreement and popularity are neutral. Finally, on average, the posts have 4.47 comments and have 13.65 shares, values that reveal the involvement of the fans / followers with the brand. When we have the correct metrics, we can analyze whether the values are good or not for the brand.

Table 47: Distribution of the interaction attributes

Interaction attributes	Mean	Standard deviation	Minimum	Median	Maximum
Likes	281.18	383.18	0.00	186.00	4571.00
WOW	0.11	0.74	0.00	0.00	12.00
Love	3.50	14.40	0.00	0.00	378.00
Haha	0.03	0.22	0.00	0.00	5.00
Total Positive Reactions	284.82	393.13	0.00	190.00	4786.00
Angry	0.01	0.17	0.00	0.00	6.00
Sad	0.00	0.07	0.00	0.00	1.00
Total Negative Reactions	0.02	0.20	0.00	0.00	7.00
Total Reactions	284.84	393.14	0.00	190.00	4787.00
Public Agreement	1.00	1.38	0.00	0.67	16.80
Popularity	1.00	1.23	0.00	0.66	21.32
Comments	4.47	7.30	0.00	2.00	93.00
Shares	13.65	16.76	0.00	9.00	291.00

4.2.3 RELATIONSHIP BETWEEN PRODUCTS ASSOCIATED TO THE POSTS WITH THE INTERACTION LEVEL

Table 48 shows the interaction of fans / followers in the posts regarding the products associated in each post. The public agreement is much larger (M = 1.23) when it has associated product than when it does not have (M = 0.66). Popularity and Public Involvement are both a bit higher if a post has product associated. Having associated products in posts, increases the metrics of the interaction, although in shares and in comments but the difference is not very accentuated. Whatever the season of the product, the metrics are not affected.

Analyzing Table 44, it is possible to verify that not having belts leads to better indicators of interaction than when the post is associated with belts. Wallets, textiles, hair articles, clothes, hats and winter articles has the same indicators. When the post is associated to other products the indicators of interaction are higher when they are associated to the posts. The exception is when the post is associated to travel. In this case the values of indicators are considerably higher.

Table 48: Interaction metrics by products associated

		Posts	Public Agreement		Popularity		Public Involvement		Reactions		Shares		Comments	
			Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Product Associated	No	753	0.66	1.01	0 .89	1.52	0.80	1.49	188.44	286.97	12.21	20.69	3.57	6.68
	Yes	1092	1.23	1.54	1.07	0.98	1.14	1.71	351.30	440.05	14.64	13.32	5.10	7.65
Season	Fall/Winter	625	1.16	1.19	1.11	1.14	1.12	1.55	331.26	338.15	15.18	15.60	4.99	6.95
	Spring/Summer	530	1.26	1.89	1.09	1.15	1.15	1.86	359.37	537.27	14.92	15.75	5.14	8.31
	Generic	690	0.65	0.94	0.83	1.34	0.78	1.49	185.53	269.03	11.28	18.23	3.49	6.67
Different Products Associated	Without Product	753	0.66	1.01	0.89	1.52	0.80	1.49	188.44	286.97	12.21	20.69	3.57	6.68
	No	976	1.25	1.56	1.04	0.98	1.11	1.62	357.29	445.62	14.24	13.41	4.95	7.25
	Yes	116	1.06	1.36	1.32	0.89	1.41	2.32	300.93	388.10	18.03	12.09	6.31	10.36
Belts	No	1802	1.01	1.39	1.00	1.24	1.01	1.65	286.98	397.33	13.67	16.91	4.50	7.37
	Yes	43	0.68	0.31	0.92	0.61	0.79	0.86	194.81	88.82	12.56	8.28	3.53	3.86
Watches	No	1804	1.00	1.39	0.99	1.24	0.99	1.64	283.51	395.41	13.57	16.86	4.44	7.34
	Yes	41	1.20	0.96	1.25	0.83	1.35	1.26	343.02	272.97	17.00	11.32	6.02	5.66
Wallets	No	1795	1.00	1.40	1.00	1.23	1.00	1.64	285.77	397.92	13.59	16.84	4.48	7.36
	Yes	50	0.88	0.47	1.14	1.01	0.96	1.11	251.26	135.23	15.58	13.73	4.28	4.96
Shoes	No	1733	0.99	1.41	0.99	1.25	0.97	1.65	281.50	400.68	13.47	17.08	4.36	7.39
	Yes	112	1.18	0.86	1.20	0.77	1.40	1.24	336.38	244.51	16.32	10.49	6.25	5.54
Bijou	No	1525	1.00	1.45	1.00	1.29	1.04	1.69	283.86	411.94	13.69	17.62	4.67	7.56
	Yes	320	1.02	1.01	0.98	0.87	0.79	1.31	289.49	287.65	13.44	11.87	3.53	5.84
Textiles	No	1716	1.02	1.42	1.01	1.26	1.02	1.67	291.25	405.36	13.75	17.16	4.57	7.47
	Yes	129	0.70	0.46	0.90	0.74	0.72	0.97	199.50	130.96	12.22	10.11	3.20	4.34
Bags	No	1610	0.86	1.10	0.90	1.19	0.85	1.44	246.10	314.53	12.33	16.22	3.82	6.46
	Yes	235	1.93	2.37	1.66	1.29	2.00	2.35	550.21	675.79	22.66	17.65	8.94	10.51

		Posts	Public Agreement		Popularity		Public Involvement		Reactions		Shares		Comments	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Night Bags	No	1697	0.97	1.35	0.99	1.27	0.98	1.65	277.41	385.27	13.55	17.26	4.41	7.38
	Yes	148	1.30	1.64	1.08	0.68	1.17	1.41	369.99	467.36	14.71	9.23	5.25	6.30
Party Bags	No	1776	1.00	1.40	1.00	1.25	1.00	1.65	285.04	399.36	13.68	17.03	4.48	7.37
	Yes	69	0.98	0.59	0.94	0.53	0.95	1.19	279.51	168.16	12.81	7.30	4.23	5.35
Sunglasses	No	1782	1.00	1.37	1.00	1.24	1.00	1.58	285.31	388.91	13.66	16.87	4.48	7.08
	Yes	63	0.95	1.76	0.98	0.98	0.99	2.71	271.32	502.07	13.37	13.43	4.44	12.15
Travel	No	1842	0.99	1.35	0.99	1.22	0.99	1.56	282.52	383.34	13.58	16.66	4.42	7.00
	Yes	3	6.00	7.19	4.10	2.30	8.86	10.48	1708.67	2048.08	56.00	31.43	39.67	46.88
Hair Art.	No	1822	1.00	1.39	1.01	1.23	1.01	1.64	285.73	394.76	13.72	16.84	4.50	7.34
	Yes	23	0.75	0.79	0.57	0.54	0.60	0.82	214.09	226.26	7.83	7.36	2.70	3.66
Clothes	No	1781	1.01	1.40	1.01	1.25	1.01	1.65	286.71	398.65	13.76	17.00	4.54	7.40
	Yes	64	0.82	0.62	0.76	0.50	0.58	0.70	232.75	176.30	10.38	6.76	2.61	3.15
Hats	No	1825	1.00	1.39	1.00	1.23	1.00	1.64	285.09	394.90	13.69	16.84	4.49	7.34
	Yes	20	0.92	0.60	0.70	0.35	0.60	0.69	261.50	170.17	9.50	4.82	2.70	3.08
Winter Art.	No	1831	1.00	1.38	1.00	1.23	1.00	1.63	283.68	393.53	13.60	16.80	4.45	7.29
	Yes	14	1.53	1.11	1.46	0.77	1.64	1.94	435.36	314.91	19.86	10.52	7.36	8.70

Note: SD = Standard deviation

4.2.4 RELATIONSHIP OF THE PRODUCTS ASSOCIATED TO THE POSTS WITH THE SENTIMENT OF THE POST

Regarding sentiment present in the posts, Table 49 show that the percentage of the products associated to posts with sentiments are generally higher than the posts with no sentiments associated.

Post sentiment is present in 67.9% of the posts where a product is associated and when verifying the season, the values are almost the same. The posts with products associated with shoes, bags, hair articles, have higher percentage of posts without present feeling than when they do not refer these products.

Additionally, posts with travel product associated have sentiment present, when compared with only 63.9% if they do not respect travel products.

Regarding Table 50, it shows that the percentage of positive sentiments presented in the posts is almost always between 50% and 60% and the neutral sentiment is between 35% and 50%.

Party Bags has the higher percentage of negative sentiments in posts. Hair articles is the product with more neutral posts.

Travel with 100% and bags with 81.70% are the products with positive sentiments in posts. Thus, if the posts have product associated or not, it has more than 50% of positive sentiments.

Both seasons have posts with positive sentiments. Negative sentiments are present in posts but with values that are no so high.

Analyzing Table 50, it is possible to verify that different products associated in the same post with more weight and positives are shoes, bags, night bags and travel. With less weight of positives are sunglasses, hair articles and clothes.

Table 49: Post sentiment present by products associated

		Post Sentiment Present					
		No		Yes		Total	
		Posts	%	Posts	%	Posts	%
Product Associated	No	314	41.70	439	58.30	753	100.00
	Yes	351	32.14	741	67.86	1092	100.00
Season	Fall/Winter	202	32.32	423	67.68	625	100.00
	Spring/Summer	169	31.89	361	68.11	530	100.00
	Generic	294	42.61	396	57.39	690	100.00
Different Products	Without Product	314	41.70	439	58.30	753	100.00
Associated	No	325	33.30	651	66.70	976	100.00
	Yes	26	22.41	90	77.59	116	100.00
Belts	No	650	36.07	1152	63.93	1802	100.00
	Yes	15	34.88	28	65.12	43	100.00
Watches	No	655	36.31	1149	63.69	1804	100.00
	Yes	10	24.39	31	75.61	41	100.00
Wallets	No	647	36.04	1148	63.96	1795	100.00
	Yes	18	36.00	32	64.00	50	100.00
Shoes	No	643	37.10	1090	62.90	1733	100.00
	Yes	22	19.64	90	80.36	112	100.00
Bijou	No	542	35.54	983	64.46	1525	100.00
	Yes	123	38.44	197	61.56	320	100.00
Textiles	No	623	36.31	1093	63.69	1716	100.00
	Yes	42	32.56	87	67.44	129	100.00
Bags	No	630	39.13	980	60.87	1610	100.00
	Yes	35	14.89	200	85.11	235	100.00
Night Bags	No	629	37.07	1068	62.93	1697	100.00
	Yes	36	24.32	112	75.68	148	100.00
Party Bags	No	644	36.26	1132	63.74	1776	100.00
	Yes	21	30.43	48	69.57	69	100.00
Sunglasses	No	639	35.86	1143	64.14	1782	100.00
	Yes	26	41.27	37	58.73	63	100.00
Travel	No	665	36.10	1177	63.90	1842	100.00
	Yes	0	0.00	3	100.00	3	100.00
Hair Art.	No	651	35.73	1171	64.27	1822	100.00
	Yes	14	60.87	9	39.13	23	100.00
Clothes	No	638	35.82	1143	64.18	1781	100.00
	Yes	27	42.19	37	57.81	64	100.00
Hats	No	658	36.05	1167	63.95	1825	100.00
	Yes	7	35.00	13	65.00	20	100.00
Winter Art.	No	660	36.05	1171	63.95	1831	100.00
	Yes	5	35.71	9	64.29	14	100.00

Table 50: Post sentiment by products associated

		Post Sentiment							
		Negative		Neutral		Positive		Total	
		Posts	%	Posts	%	Posts	%	Posts	%
Product Associated	No	21	2.79	345	45.82	387	51.39	753	100.00
	Yes	46	4.21	399	36.54	647	59.25	1092	100.00
Season	Fall/Winter	25	4.00	219	35.04	381	60.96	625	100.00
	Spring/Summer	24	4.53	202	38.11	304	57.36	530	100.00
	Generic	18	2.61	323	46.81	349	50.58	690	100.00
Different Products	Without Product	21	2.79	345	45.82	387	51.39	753	100.00
Associated	No	45	4.61	367	37.60	564	57.79	976	100.00
	Yes	1	0.86	32	27.59	83	71.55	116	100.00
Belts	No	64	3.55	727	40.34	1011	56.10	1802	100.00
	Yes	3	6.98	17	39.53	23	53.49	43	100.00
Watches	No	65	3.60	732	40.58	1007	55.82	1804	100.00
	Yes	2	4.88	12	29.27	27	65.85	41	100.00
Wallets	No	65	3.62	724	40.33	1006	56.04	1795	100.00
	Yes	2	4.00	20	40.00	28	56.00	50	100.00
Shoes	No	65	3.75	716	41.32	952	54.93	1733	100.00
	Yes	2	1.79	28	25.00	82	73.21	112	100.00
Bijou	No	59	3.87	606	39.74	860	56.39	1525	100.00
	Yes	8	2.50	138	43.13	174	54.38	320	100.00
Textiles	No	63	3.67	695	40.50	958	55.83	1716	100.00
	Yes	4	3.10	49	37.98	76	58.91	129	100.00
Bags	No	65	4.04	703	43.66	842	52.30	1610	100.00
	Yes	2	0.85	41	17.45	192	81.70	235	100.00
Night Bags	No	59	3.48	701	41.31	937	55.22	1697	100.00
	Yes	8	5.41	43	29.05	97	65.54	148	100.00
Party Bags	No	61	3.43	720	40.54	995	56.02	1776	100.00
	Yes	6	8.70	24	34.78	39	56.52	69	100.00
Sunglasses	No	63	3.54	712	39.96	1007	56.51	1782	100.00
	Yes	4	6.35	32	50.79	27	42.86	63	100.00
Travel	No	67	3.64	744	40.39	1031	55.97	1842	100.00
	Yes	0	0.00	0	0.00	3	100.00	3	100.00
Hair Art.	No	66	3.62	729	40.01	1027	56.37	1822	100.00
	Yes	1	4.35	15	65.22	7	30.43	23	100.00
Clothes	No	63	3.54	710	39.87	1008	56.60	1781	100.00
	Yes	4	6.25	34	53.13	26	40.63	64	100.00
Hats	No	66	3.62	735	40.27	1024	56.11	1825	100.00
	Yes	1	5.00	9	45.00	10	50.00	20	100.00
Winter Art.	No	67	3.66	739	40.36	1025	55.98	1831	100.00
	Yes	0	0.00	5	35.71	9	64.29	14	100.00

4.2.5 PREDICTIVE MODEL OF THE POSITIVE SENTIMENT ASSOCIATED TO THE POST

The evaluation metrics of the models obtained with decision trees (CART and C5 algorithms) and artificial neural networks are presented in Tables 47 and 48. Regarding the CART algorithm Table 51 shows the evaluation metrics. The best model in the test sample for CART, was the model A (accuracy=63.05%, sensibility 71.99% and specificity is 52.29%. Still, model A has the best specificity.

Table 51: Evaluation metrics for CART algorithm

	Model A: CART		**Model B: CART boosting**		**Model C: CART bagging (voting)**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Accuracy	61.62%	63.05%	66.01%	57.44%	67.43%	57.09%
Sensibility	68.64%	71.99%	84.87%	78.83%	85.14%	77.20%
Specificity	52.29%	52.65%	40.95%	32.58%	43.88%	33.71%
AUC	0.637	0.651	0.720	0.596	0.710	0.610

Regarding C5 algorithm, Table 52 shows the performance metrics, being the best model in the test sample the model D (accuracy=63.4%, sensibility 76.2% and specificity is slightly below 50% but, by far, this model has the best specificity).

Table 52: Evaluation metrics for C5 algorithm

	Model D: C5 favour accuracy		**Model E: C5 favour generality**		**Model F: C5 boosting favour accuracy**		**Model F: C5 boosting favour generality**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Accuracy	64.36%	63.57%	60.52%	59.19%	60.83%	59.37%	60.52%	59.19%
Sensibility	75.79%	76.22%	83.91%	83.39%	84.46%	84.04%	83.91%	83.39%
Specificity	49.18%	48.86%	29.43%	31.06%	29.43%	30.68%	29.43%	31.06%
AUC	0.656	0.655	0.562	0.570	0.567	0.572	0.562	0.570

Finally, Table 53 shows the metrics for the artificial neural networks, being the Model H the best in the test sample (accuracy=57.6%, sensibility=66.8% and specificity=47.0%).

Table 53: Evaluation metrics for ANN algorithm

	Model H: ANN		Model I: ANN boosting		Model J: ANN bagging	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Accuracy	61.85%	57.62%	97.72%	57.27%	95.60%	57.27%
Sensibility	68.50%	66.78%	97.11%	63.84%	95.74%	64.50%
Specificity	53.02%	46.97%	98.54%	49.62%	95.43%	48.86%
AUC	0.653	0.602	0.977	0.595	0.951	0.554

To compare the models that perform better with the three different algorithms, Figure 18 shows the ROC (Receiver Operating Characteristic) curve and consequently the AUC (Area Under the Curve). It is also possible to conclude that de C5 model is the one that better predicts the positive sentiment associate to the post. Figure 18 shows the rule set associated to the decision tree and Figure 19 the most important predictors. Post reason is the most important predictor (importance=0.350), followed by the number of products associated (importance=0.272) and by the month of the post (importance=0.153).

Figure 18: ROC of the three predictive models

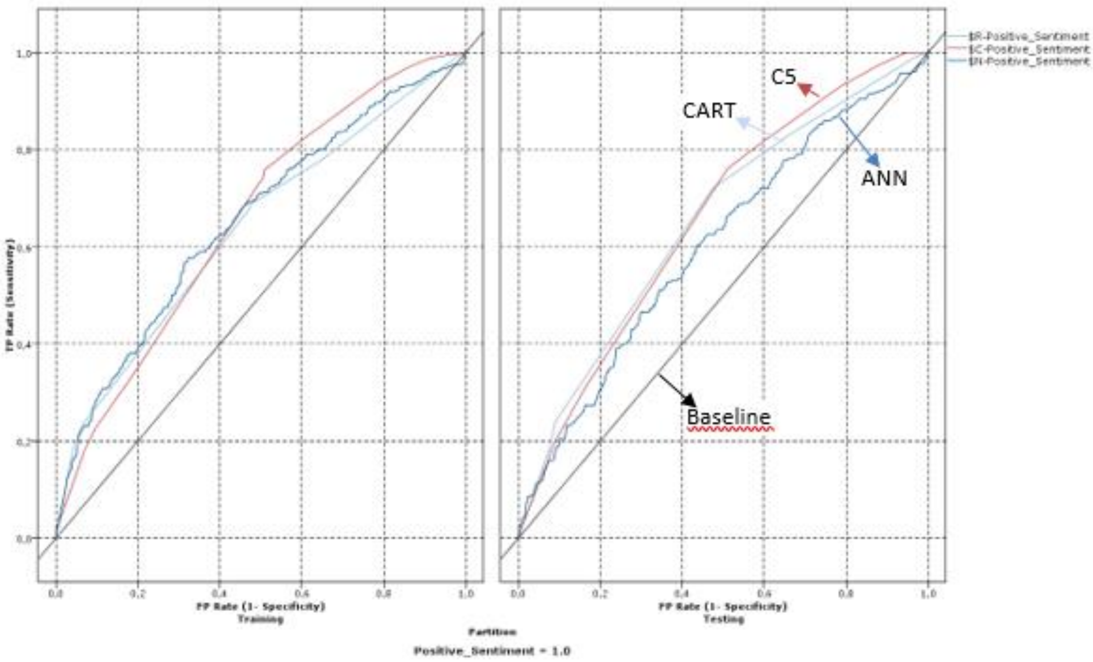
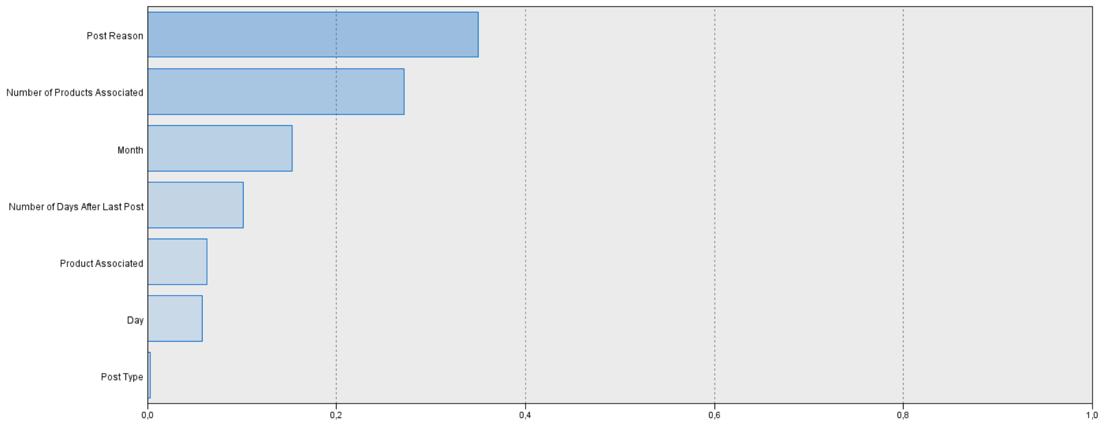


Figure 19: Rule set for the best predictive model of the positive sentiment (C5 model)

- Number of Products Associated <= 1 [Mode: 1] (1.639)
  - Post Reason in [Announcement Blog] [Mode: 1] ⇨ **1,0** (184,362; 0,623)
  - Post Reason in [Campaign Contest] [Mode: 0] ⇨ **0,0** (269,452; 0,555)
  - Post Reason in [Job Offer] [Mode: 0] (21,273)
    - Number of Days After Last Post <= 0 [Mode: 0] ⇨ **0,0** (18,091; 0,885)
    - Number of Days After Last Post > 0 [Mode: 1] ⇨ **1,0** (3,182; 0,673)
  - Post Reason in [New Store] [Mode: 0] (109,402)
    - Post Type = 1,000 [Mode: 0] ⇨ **0,0** (0)
    - Post Type = 2,000 [Mode: 0] ⇨ **0,0** (102,393; 0,628)
    - Post Type = 3,000 [Mode: 1] ⇨ **1,0** (7,009; 0,857)
  - Post Reason in [Newspaper] [Mode: 0] (16,208)
    - Number of Days After Last Post <= 1 [Mode: 1] ⇨ **1,0** (7,178; 0,577)
    - Number of Days After Last Post > 1 [Mode: 0] ⇨ **0,0** (9,03; 0,997)
  - Post Reason in [Sales] [Mode: 1] (1.038,303)
    - Month in [Jan Jul] [Mode: 0] ⇨ **0,0** (145,267; 0,537)
    - Month in [Feb Mar Apr May Aug Sep Oct Nov] [Mode: 1] ⇨ **1,0** (694,136; 0,622)
    - Month in [Jun] [Mode: 0] (73,267)
      - Product Associated = 1,000 [Mode: 0] ⇨ **0,0** (63,633; 0,676)
      - Product Associated = 0,000 [Mode: 1] ⇨ **1,0** (9,633; 0,792)
    - Month in [Dec] [Mode: 1] (125,633)
      - Day in [1 2 3 6 8 10 13 15 17 18 20 22 24 26 27 28 29 31] [Mode: 1] ⇨ **1,0** (67,633; 0,704)
      - Day in [4] [Mode: 1] (5)
        - Product Associated = 1,000 [Mode: 1] ⇨ **1,0** (3; 1,0)
        - Product Associated = 0,000 [Mode: 0] ⇨ **0,0** (2; 1,0)
      - Day in [5 7 11 12 14 19 23 25 30] [Mode: 0] ⇨ **0,0** (35; 0,8)
      - Day in [9] [Mode: 1] (6)
        - Product Associated = 1,000 [Mode: 0] ⇨ **0,0** (3; 1,0)
        - Product Associated = 0,000 [Mode: 1] ⇨ **1,0** (3; 1,0)
      - Day in [16] [Mode: 1] (6)
        - Product Associated = 1,000 [Mode: 0] ⇨ **0,0** (2; 1,0)
        - Product Associated = 0,000 [Mode: 1] ⇨ **1,0** (4; 0,75)
      - Day in [21] [Mode: 0] (6)
        - Number of Days After Last Post <= 0 [Mode: 1] ⇨ **1,0** (2; 1,0)
        - Number of Days After Last Post > 0 [Mode: 0] ⇨ **0,0** (4; 1,0)
- Number of Products Associated > 1 [Mode: 1] ⇨ **1,0** (206; 0,748)

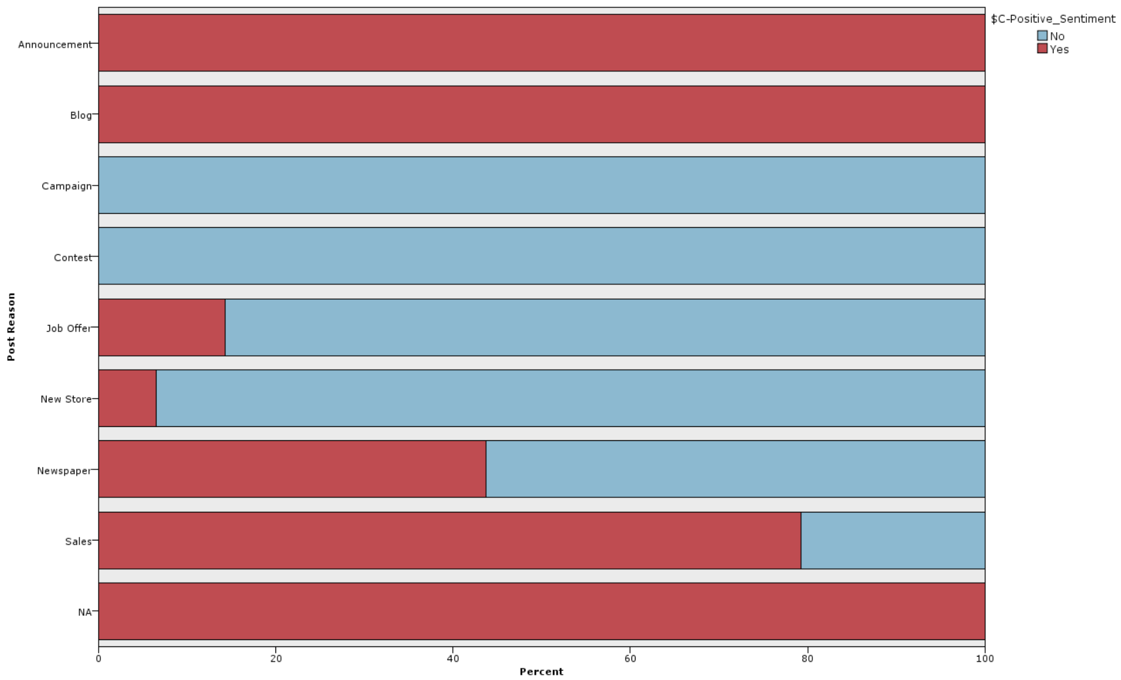
Figure 20: Predictors importance (target = positive sentiment)



In order to better understand the model, a sensibility analysis is performed. In this analysis, the relationship between the predicted class (positive sentiment) and the more important predictors are characterized. Figures 20 to 22 show these relationships. The red color is positive sentiment and blue is negative sentiment.

Clearly the reasons, announcement, blog and NA are associated with positive sentiments (100%), also posts motivated by sales have a high percentage (almost 80%) of a positive sentiment. Thus, the posts motivated by campaigns and contest have clearly associated a non-positive sentiment (100%) and also the reasons job offer and new store tend to lead to a non-positive sentiment. Finally, the newspaper reason divides the feelings, with even more than half in which the model predicts a non-positive sentiment.

Figure 21: Distribution of the predicted sentiment by post reason



In Figure 21 clearly January, June and July are the months with a non-positive sentiment prediction. All the others lead to a positive sentiment.

Figure 22: Distribution of the predicted sentiment by month of the year

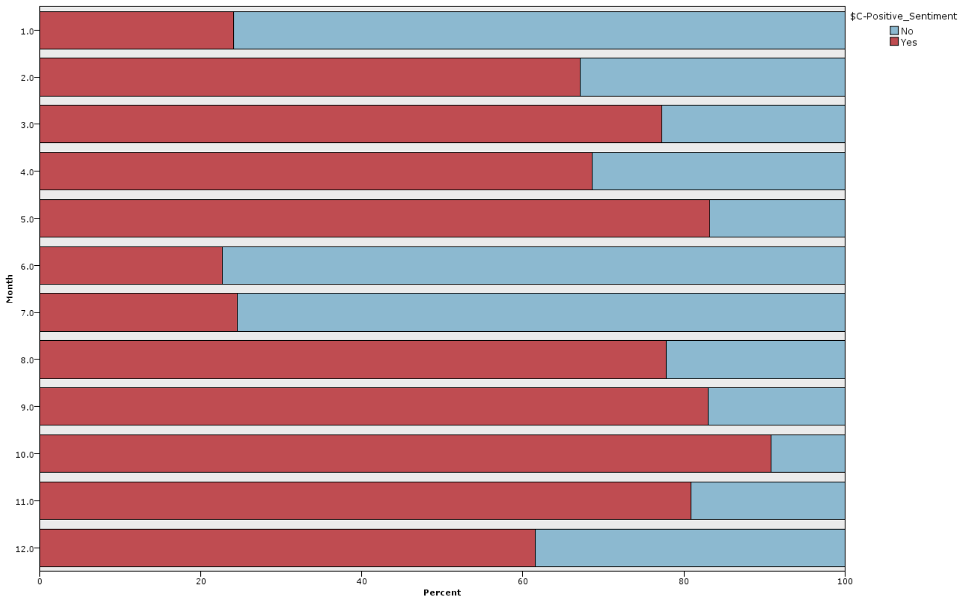
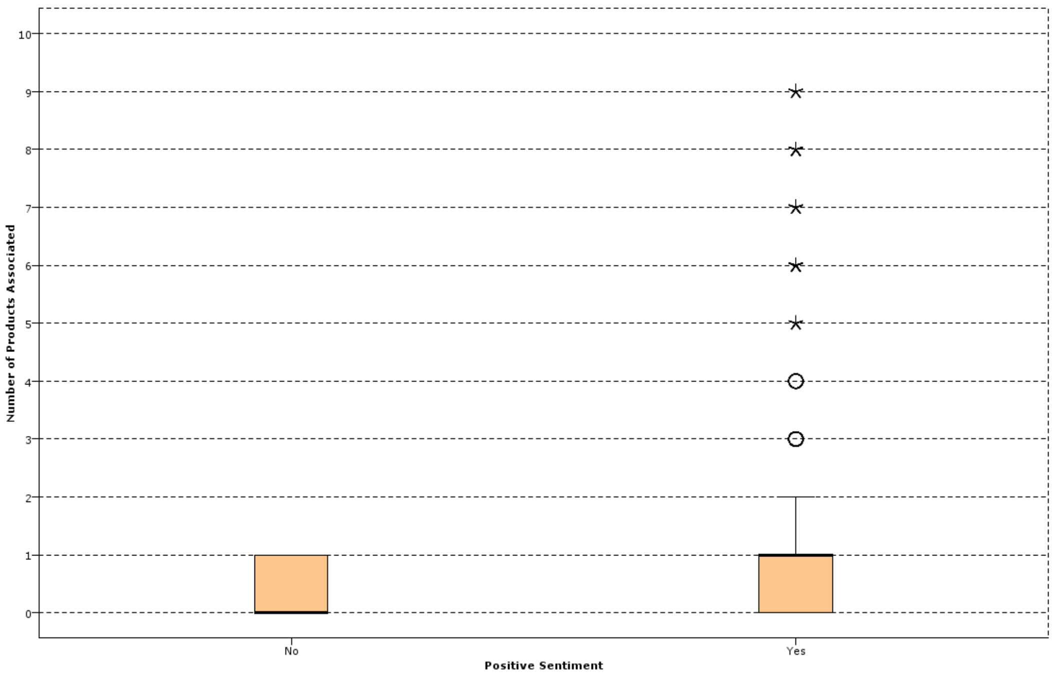


Figure 23 shows that posts with fewer associated products tend to lead to non-positive sentiment. As show, half of the posts without associated product have a prediction of non-positive feeling. Instead, for the positive feeling it is verified that half of the posts with one or more associated products lead to a positive sentiment prediction.

Figure 23: Distribution of the number of products associated by predicted sentiment



4.3 DISCUSSION

The results of this research give important knowledge to the managers and, also, to academics. For Parfois managers some results are very interesting, namely:

- More than 50% of the comments about accessories, clothing, editions and wallets have a positive sentiment. In the other hand, negative sentiments are more relevant in comments about animals and customer service. It happens as Parfois have some products with real fur. This is always a sensible matter, that involves the animal's advocates. Is also known that people tend to express their displeasure in SMP, as the visualization of the comments reaches more people;

- Also, more than 50% of the comments made in German or Polish have positive sentiments. Romanian, Spanish and Portuguese languages have the higher percentage of comments with negative sentiments. Exists some empirical evidences, that prove some cultures are more suitable to negativism and so, more suitable to create negative content rather than positive one. Moreover, this characteristic is associated to Latin countries. This is the case of Portugal and Spanish;

- The number of the posts are not the same all years. These could happen because is not a proper SM strategy or even the use of another SMP, case of Tweeter and Instagram. Moreover, in many times it is posted more than one post in one day and is not posted every day. The reason it to avoid being considered spam and, thus, affect the reputation of the brand;

- Posts with photo have, by far, the higher percentage and when checking the post reason, the higher percentage is sales. These results mean that Parfois uses photos to show products in order to increase their sales. For selling the strategy is use posts with photos. Moreover, clearly the reasons, announcement, blog, NA and sales are associated with positive sentiments. Thus, the posts motivated by campaigns, contest, job offer and new store have clearly associated a non-positive;

- Moreover, bijou and bags are the product categories with more posts. The other products have almost an equality percentage of posts mentions. This result reflects the two categories with more percentage of sales. Thus, the brand is not using the posts in Facebook to advertise products with less percentage of sales, so it could have a chance to increase the sale of that products;

- Posts with fewer associated products tend to lead to non-positive sentiment. For example, half of the posts without associated product have a prediction of non-positive feeling. Already for the positive feeling it is verified that half of the posts with one or more associated products lead to predict a positive feeling. Thus, must Parfois post more posts with different products?

- Certain products have less interaction than posts that do not refer to this product as for example belts, wallets, textiles, hair articles, clothes, hats and winter articles. Posts with products associated like shoes or bags have a higher percentage of posts with sentiment present than when they do not refer to these products. Additionally, posts with associated travel product have sentiment present in all posts, when compared with only 63.9% if they do not respect travel;

- Regarding interaction metrics, on average, the public involvement is low, the public agreement is neutral as it is the popularity. Although, on average, the posts have 4.47 comments, 13.65 shares and 281 likes. These values reveal the involvement of the fans / followers with the brand and almost always a fan / follower interacts with the brand. Moreover, 56% of the posts have a positive sentiment associated and only 3.6% have a negative sentiment, meaning that Parfois has a bit more than the half value of total posts with positive sentiments and the fans / follower’s reactions are also positive.

In this way Parfois can select different products for the post according to needs and recent trends. If the last posts have associated a negative feeling, then a new post with the characteristics that lead to a positive feeling must have resisted to reverse the tendency. All the results can help the brand, think, design new products, prepare new seasons and build the brand having in mind fans / follower’s comments and sentiments.

The results of the research are important but they should be shared in different perspectives to the different departments. For marketing department is important measures like popularity, public involvement or public agreement as these measures can give percentages of the involvement of the fans / followers with the brand. Regarding the sentiments associated to products or features must draw attention of the design team since it conveys the wishes or complaints of the fans / followers about products or features. Complaints about customer service must be analyzed by the e-commerce department. Furthermore, financial department must analyze the association of products posted and their correspondent sales, as they can measure the interval before and after the post. So, each different department can analyze the results as one piece of a big puzzle.

Thus, for this research, some of them are more important. As concluded by Bose (2011), the integration of external data into organizational BI, can support various processes in sales, customer service and operations, just to name some of them. Thus, KPIs like popularity, public involvement, public agreement, number of likes in a post, number of shares, number of comments made in a certain post and the percentage of positive sentiments and must have a dashboard where they can be monitored.

Finally, to help to understand the results, some questions were asked to Engineer Ricardo Luis Moura about the importance of SA and inclusion of the results in the organizational BI. Engineer Moura has a degree in Systems and Informatics Engineer, degree that was given by University of Minho and has a large experience in Business Intelligence area, and has the roll of Business Intelligence Manager in the company where he works.

What is the value added by Sentiment Analysis?

Knowing Parfois, has he already worked in the company, Engineer Moura says: “it makes sense that in a company like Parfois that sells in the base of emotion and impulse, we find a mixture between segment and moment”.

Regarding BI, what analyses are important to make?

Is important to study the expressions, state of mind, which key words were used, which is the commentary object, what was the origin of the comment or even the social context of the person (those who can afford more tend to be happier).

In which manner can the positive/negative sentiments help?

When you buy by impulse, the state of mind has an impact estimated between 80% and 90%. For example, in Dubai exists a denominated concept that is happiness meter. People answer to inquires in apps about locations and how they are / were happy on those locations. In that way, they can increase all type of offers, whether they are playful, commercial or cultural in that places. One of Dubai's strategic aims is to be the country with the happiest population in the world.

What kind of indicators are important and should be analyzed?

Some indicators that should be analyzed, if possible, are: degree of happiness; positive words / negative words; and Emojis. It is important to emphasize that a positive word and a negative word will give a neutral result.

In a future conversation with the Marketing, E-commerce directors, or even with the CEO of the brand, the results of this research should be analyzed and inserted into the strategy for SM of Parfois. Additionally, for some metrics to be inserted into BI and can be monitored for example using dashboards, you must set targets for key metrics and identify possible paths for deviations.

Moreover, as Moro et al. (2015: p. 3350) state this research" … may be enriched with other context features (e.g., if the product is being advertised elsewhere) for tuning its performance. Also, text mining methods could be employed to the content for extracting additional knowledge. Finally, text mining the comments of each post for user sentiment analysis could reveal the feelings each post is generating.".

Finally, the results of this research can help a brand to understand better the fans / followers and get an important CI.

5 CONCLUSION

This chapter aims to present the conclusions regarding the overall research, namely, the work carried out since the bibliographic collection to the work made in methodology, the main difficulties experienced, the limitations of the project and, lastly, the proposals and future implementations.

5.1 SUMMARY

This research objectives were to make a sentiment analysis in a retail brand, more specifically in a Portuguese fashion brand, Parfois. Along with the sentiment analysis it was performed social media analysis using metrics, focusing engagement, using content generated by users in SMP, more in particular Facebook. For this purpose, sentiment polarity was extracted through TM techniques of 1.845 posts and 8.256 comments. In this research, it was used open-source tools to create the extraction tool and to translate comments to English. Moreover, it was used CRISP-DM methodology, as it focuses on business approach. To make the TM it was used IBM SPSS Modeler, as to create the decision tree and neural networks models.

Along with the results, other objective was the creation of an understanding about SM and the benefits of SMA to the business. This research set an overall approach, so it can use user generated content from Facebook and from public sources to extract data and integrate it in organizational BI. In this regard, the problems underlying this type of analysis were also described and what techniques are used to make a correct analysis.

.

5.2 CONTRIBUTIONS

The results of this research gave the knowledge about the comments and reactions made by fans / followers of Parfois Facebook page and how transform it into meaningful information, that can be very important for several departments. This engagement between fans / followers with the brand gave the information about what are the words most commonly used in the comments, the sentiments associated to the posts and products associated to the posts. This SMA can give a set of data which can definitively help in brand management as it is possible to verify the sentiment associated to a product and what type of comments are being generated by fans / followers.

The contribution of this research relies on the use of unstructured data from Facebook and have the perception of the sentiments and the interactions of the fans / followers of the brand. As so, the present research fills the void in SA in retail, giving managers the knowledge to make better decisions and, thus, contributing to brand building.

Although, for the brand is important to know the sentiments of fans / followers about the brand, it is also important to know the sentiments of fans / followers towards the products posted in Parfois Facebook page, giving the brand specific information that can help when the brand is preparing a new season. With all the analyses, it is possible to crosscheck each category percentage sales and adapt the posts in order to increase sales of a certain category or product.

Another contribution to the brand managers is knowing what types of posts decrease the sentiments towards the brand, and avoid them or after one that decrease, make another one that will increase the sentiment. Also, it is possible to prepare a new approach to make posts in a way that will increase all engagement measures like the reactions or even increase the number of comments, increasing the positive sentiments too.

This research gives a contribution to the literature in several ways. Firstly, it provides a SA process applied to retail, contributing to several departments of an organization, so it can help to increase competitive advantage and help organizations decision makers. Additionally, due the lack of literature, this research gives a practical point of view about SA and its application in a fashion brand.

5.3 LIMITATIONS

Despite all the work done in this research, it is important to emphasize some limitations encountered, such as:

- Despite the sample size, data can be insufficient to identify all the patterns;
- Although the various languages collected from Parfois Facebook, it would improve the analysis having a higher number of comments from diversified geography (Parfois Facebook pages created by partners around the world);
- The subjectivity when applying TM techniques is another limitation, as for example in the dictionary definition. Another limitation is the natural flaw when analyzing word with multiple meanings, as they can be associated to a positive sentiment when it was referring to another. In this research it was minimized when IBM SPSS Modeler was chosen as the tool to apply the TM and SA techniques.

5.4 FUTURE WORK

For future work, it suggests to the brand to create a structure in organizational BI in order to have external data in it. For external data it is comprehend the analysis of official Facebook Page, Parfois pages in other countries and compare the results. Another analysis for external data is the extraction and analysis of other SMPs where Parfois has an official page like Twitter and Instagram.

Joining all these external data and correspondent analysis to internal data, will give a new type of information based in SA, in this particular case. The goal is to have a new set of reports that will give a whole new meaningful set of data for the decision makers. Internally in Parfois the decision makers who will use the reports are the Online Customer Service team, Product manager and Marketing department.

The data that will be incorporated in the reports is the data coming from the analysis of proactive follower activity and corresponding sentiment (in concrete: Analysis of Parfois posts and corresponding comments, likes, shares). The teams require answers to questions as:

- Is a user-generated post relevant for our business activity?
- What is the content of a user-generated post? / Which topics were mentioned?

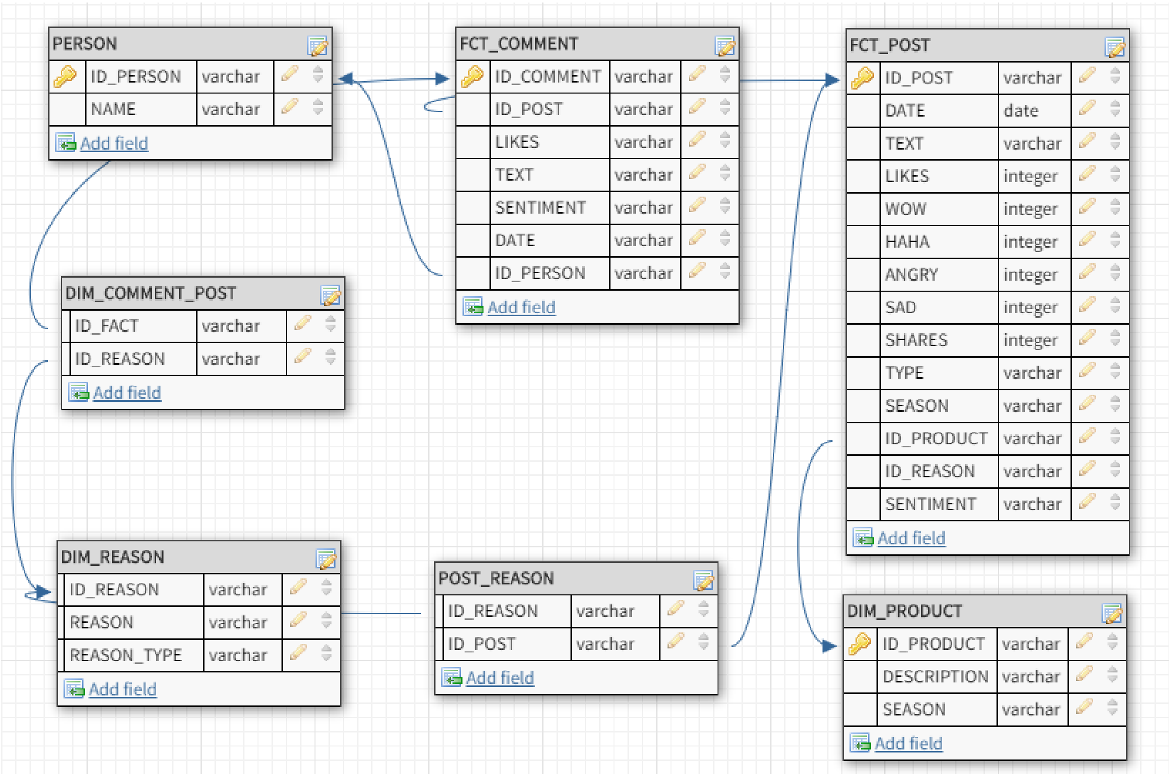
- What is the sentiment of the post?
- Do other users share the same opinion?
- What do other users think of his post?
- How controversial is this user post?
- How do we have to react in the best way to the user's post?

To prepare the system, it will be necessary to create the following:

- Dimension tables : Dim_reasom, Dim_comment_post, Dim_product ;
- Fact Tables: Fct_comment, Fct_post;
- Granularity: Comment (Fct_comment)
- Granularity: Post (Fct_post)
- Measurement: likes, shares, comments, sentiment

Figure 24 show the suggestion of the dimension and fact tables to create and use for the reporting.

Figure 24: Suggestion of dimension and fact tables



6 REFERENCES

Afetian S., 2013, Rise of The Interest-Based Network from: http://www.cmo.com/features/articles/2013/11/27/rise_of_the_interest.html#gs.2x0D210 accessed at 28/09/2017

Agarwal A., B. Xie, I. Vovsha, O.Rambow, and R. Passonneau, “Sentiment Analysis of Twitter Data in Workshop on Language in Social Media”, In Proc. of LSM 2011 - Workshop on Languages in Social Media, 23 June, pp. 30-38, 2011.

Ahmad et al., 2015, The Impact of Social Media on Fashion Industry: Empirical Investigation from Karachiites, Journal of Resources Development and Management An International Peer-reviewed Journal Vol.7, 2015

Ananda, A.S., Hernández-García, Á, and Lamberti, L. (2015), “Social media marketing in Italian luxury fashion”. 5th Annual International Workshop on Luxury Retail, Operations and Supply Chain Management in Milan, Italy, 25-27 May 2015, Politecnico di Milano, Milan

Angella Jiyoung Kim & Eunju Ko (2010) Impacts of Luxury Fashion Brand’s Social Media Marketing on Customer Relationship and Purchase Intention, Journal of Global Fashion Marketing, 1:3, 164-171, DOI:10.1080/20932685.2010.10593068

Arboleda, Velásquez, Jiménez, (2017), A Proposal for Brand Analysis with Opinion Mining, E-Business - State of the Art of ICT Based Challenges and Solutions, http://dx.doi.org/10.5772/66567

Artha S., Ananda, Hernandez-García A., Lamberti L., 2015, Fashion in Actions on Social Media – Spanish SME Fashion Brands Case Studies, 2nd European Conference on Social Media (ECSM 2015), At Porto, Portugal, DOI: 10.13140/RG.2.1.2774.7686

Asur, S., Huberman B. (2010) Predicting the Future with Social Media, DOI: 10.1109/WI-IAT.2010.63

Ayoubloo, M. K., Azamathulla, H. Md., Jabbari, E. & Zanganeh, M. (2011). Predictive model-based for the critical submergence of horizontal intakes in open channel flows with different clearance bottoms using CART, ANN and linear regression approaches. Expert Systems with Applications. 38 (6). pp. 10114-10123. doi: 10.1016/j.eswa.2011.02.073.

Baldwin, A., 2015, These 8 types of social media should be a part of your marketing strategy online: http://blog.near-me.com/these-8-types-of-social-media-should-be-a-part-of-your-marketing -strategy, accessed at 14/07/2017

Bambauer-Sachse, S., Mangold, S. (2011). Brand equity dilution through negative online word-of-mouth communication. Journal of Retailing and Consumer Services, 18(1), pp. 38–45. DOI: http://dx.doi.org/10.1016/j.jretconser.2010.09.003

Bell K., 2015, Everything we know so far about Facebook's new reactions, from: http://mashable.com/2015/10/08/facebook-reactions-how-to/#SfzE0w0HDaq4

Berlanga R., Garcia-Moya L., Nebot V., Aramburu M. J., Sanz I., and M. Llid´o D., “Slod-bi: An open data infrastructure for enabling socialbusiness intelligence,” International Journal of Data Warehousing andMining (IJDWM), vol. 11, no. 4, pp. 1–28, 2015

Beuker, R., Abbing, E. R. (2010). Two Faces of Social Media: Brand Communication and Brand Research. Design Management Review, 21(1), pp. 54–60.

Bhadane, C., Dalal, H., & Doshi, H. (2015). Sentiment Analysis: Measuring Opinions. Procedia Computer Science, 45(0), 808-814. doi: http://dx.doi.org/10.1016/j.procs.2015.03.159

Bose, Ranjit (2011). "Discovering Business Intelligence from the Subjective Web Data". In: International Journal of Business Intelligence Research 2(4), pp. 1–16. issn: 1947-3591. doi: 10.4018/jbir.2011100101.

Boyd & Ellison Social Network Sites: Definition, History, and Scholarship (2007), Journal of Computer-Mediated Communication, Vol. 13, 13/2: 210–230

Brandtzæg, P. B., Heim, J. (2009). Why people use social networking sites. In: A. A. Ozok, P. Zaphiris (Eds.), Online communities and social computing. Springer Berlin Heidelberg, pp. 143–152.

Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. (1984). Classification and Regression Trees. Wadsworth: Taylor & Francis.

Brezigar-Masten, A. & Masten, I. (2012). CART-based selection of bankruptcy predictors for the logit model. Expert Systems with Applications. 39 (11). pp. 10153-10159. doi: 10.1016/j.eswa.2012.02.125.

Bruhn, M., Schoenmueller, V., Schäfer, D. B. (2012). Are social media replacing traditional media in terms of brand equity creation? Management Research Review, 35(9), pp.770–790.

Calder, B. J., Malthouse, E. C., Schaedel, U. (2009). An experimental study of the relationship between online engagement and advertising effectiveness. Journal of Interactive Marketing, 23(4), pp. 321–331.

Chaffey, D. (2017), Global social media research summary 2017 from: http://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research

Chapman, P. & Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). CRISP-DM 1.0 - Step-by-step data mining guide. USA: SPSS Inc.

Constantinides, E., Romero, C. L., and Boria, M. A. G. 2009. "Social Media: A New Frontier for Retailers?" in European Retail Research. Springer, pp. 1-28.

Curtis, June 2017, 10 Types of Social Media and How Each Can Benefit Your Business, from: https://blog.hootsuite.com/types-of-social-media/

Culnan, M., et al (2010) How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value, MIS Quarterly Executive Vol. 9 No. 4 / Dec 2010 Tabela esta na página 245

D'Andrea, et al. (2015) Approaches, Tools and Applications for Sentiment Analysis Implementation, International Journal of Computer Applications (0975 – 8887) Volume 125 – No.3, September 2015

Delen, D., Kuzey, C. & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. Expert Systems with Applications. 40 (10). pp. 3970-3983. doi: 10.1016/j.eswa.2013.01.012.

Dey L., Haque S., Khurdiya A., and Shroff G., “Acquiring Competitive Intelligence from Social Media”, In Proc. of the 2011 Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data, September 2011.

Dinter B., Lorenz A., “Social Business Intelligence: a Literature Review and Research Agenda,” 2012

Dongi Saothawan, 2014, Consumer Social Media Analytics in retail -Prerequisites, business value and application for Social Media Marketing

Etlinger, S., Li, C., & Groopman, J. (2013). Social data intelligence: Integrating social and enterprise data for competitive advantage. (). San Mateo, CA: Altimeter Group.

Fan, Weiguo & Gordon, Michael D. (2014). The Power of Social Media Analytics.Association for Computing Machinery Communications of the ACM,57(6),74 DOI: 10.1145/2602574

Fisher, Journal of Database Marketing & Customer Strategy Management, 2009, Volume 16, Number 3

Fontein, Dara, 2016, The Top 26 Social Media KPIs Marketers Can’t Ignore from: https://blog.hootsuite.com/social-media-kpis-key-performance-indicators/ accessed at 28/09/2017

Fuchs, C (2012) Internet and surveillance. New York, USA: Routledge

Gallinucci, E., Golfarelli, M. and Rizzi, S., “Meta-Stars: Multidimensional Modeling for Social Business Intelligence”. In: Proceedings of the 16th International Workshop on Data warehousing and OLAP – DOLAP, p. 11-18, 2013.

Gilasgar M., Mirabolghasemi M., Iahad N. A., (2016), Exploring the Potential of Social Business Intelligence in Healthcare, Journal of Information Systems Research and Innovation, Volume 9, Issue 2, pp. 1-6

Google Trends, online https://trends.google.pt/trends/ retrieved in 28/01/2017

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

He, H. & Garcia, E. A. (2009). Learning from Imbalanced Data. IEEE Transactions on Knowledge and Data Engineering. 21 (9). pp. 1263-1284.

He, W., J. Shen, X. Tian, Y. Li, V. Akula, G. Yan, and R. Tao, "Gaining Competitive Intelligence from Social Media Data", Industrial Management & Data Systems, vol. 115, no. 9, pp. 1622-1636, 2015. Doi:10.1108/IMDS-03-2015-0098.

He, W., Wu, H., Yan, G., Akula, V. and Shen, J. (2015), “A novel social media competitive analytics framework with sentiment benchmarks”, Information & Management, Vol. 52 No. 7, pp. 801–812

Heijnen, J., de Reuver, M., Bouwman, H., Warnier, M., & Horlings, H. (2013). Social media data relevant for measuring key performance indicators? A content analysis approach. Co-created effective, agile, and trusted eServices (pp. 74-84) Springer.

Hendricks, D. 2013, Complete History of Social Media: Then And Now, from: https://smallbiztrends.com/2013/05/the-complete-history-of-social-media-infographic.html, accessed at 28/09/2017

Hennig-Thurau, T., et al (2010) The Impact of New Media on Customer Relationships, in Journal of Service Research, doi: 10.1177/1094670510375460

Hogenboom, Alexander et al. (2013): Exploiting Emoticons in Sentiment Analysis. Erasmus University, Rotterdam. Available online at http://eprints.eemcs.utwente.nl/23268/01/ sac13-senticon.pdf

Hu, Wei (2013), "The impact of the knowledge sharing in social media on consumer behaviour", The Thirteen International Conference on Electronic Business, Vol. 1 No. 1, pp. 71-102.

IBM. (2016a). IBM SPSS Modeler 17 Modeling Nodes. Retrieved July 26, 2016, from IBM Web site: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/17.0/en/ ModelerModelingNodes.pdf

Ihsan M Y, Azlee Z, Rozita J, Megat S A M A, Rahimi B, Abu H A H, Zairi I R.Comparison between cascade forward and multi-layer perceptron neural networks for NARX functional electrical stimulation (FES)-based muscle model.International Journal on Advanced Science, Engineering and Information Technology, 2017, 7(1):215-221

Internet live stats (2017), Twitter Usage Statistics, Retrieved May 26, 2017 from http://www.internetlivestats.com/twitter-statistics/

Investopedia, Key Performance Indicators – KPI, online at: http://www.investopedia.com/ terms/k/kpi.asp, accessed at 11/08/2017

Jocelyn Sietsma & Robert J.F.Dow (1991), Creating artificial neural networks that generalize, Neural Networks, Volume 4, Issue 1, 1991, Pages 67-79, https://doi.org/10.1016/0893-6080(91)90033-2

Jonathan Taylor, February 22, 2017 What is a KPI, Metric or Measure? https://www.klipfolio.com/blog/kpi-metric-measure

Kallas, P. 2017, Top 15 Most Popular Social Networking Sites and Apps, from https://www.dreamgrow.com/top-15-most-popular-social-networking-sites/ accessed at 28/09/2017

Kiezmann et al. (2011): Social Media? Get Serious! Understanding the Functional Building Blocks of Social Media. In Business Horizons 2011 (54), pp. 241–251. Available online at http://www.academia.edu/959458/Social_Media_Get_Serious_Understanding_the_ Functional_Building_Blocks_of_Social_Media

Kim A. J. & Ko E. (2010) Impacts of Luxury Fashion Brand's Social Media Marketing on Customer Relationship and Purchase Intention, Journal of Global Fashion Marketing, 1:3, 164-171, DOI:10.1080/20932685.2010.10593068

Kiron, D., Perguson, R., Boucher & Prentice, Pamela K. (2013). From Valueto Vision: Reimagining the Possible with Data Analytics. MIT Sloan Management Review

Konsta K., Plomaritou E., 2012, Key Performance Indicators (KPIs) and Shipping Companies Performance Evaluation: The Case of Greek Tanker Shipping Companies, International Journal of Business and Management Vol. 7, No. 10; May 2012

Socialbakers, December 2016 Social Marketing Report Portugal https://www.socialbakers.com/resources/reports/portugal/2016/december/ retrieved in 28/01/2017

Facebook, 2017, https://www.facebook.com/help/1625715647696553

Li, H., Sun, J. & Wu, J. (2010). Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. Expert Systems with Applications. 37. pp.5895-5904. doi: 10.1016/j.eswa.2010.02.016.

Lipika Dey, Sk. Mirajul Haque, Arpit Khurdiya, Gautam Shroff (2011) Proceedings of the 2011 Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data Article No. 3, Beijing, China — September 17 - 17, 2011

Litvin, S., Goldsmith, R. and Pan, B. (2008), "Electronic word-of-mouth in hospitality and tourism management", Tourism Management, Vol. 29 No. 1, pp. 458-468.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

Liu et al., (2007), ARSA: a sentiment-aware model for predicting sales performance using blogs. DOI: 10.1145/1277741.1277845 · Source: DBLP Conference: SIGIR 2007: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, The Netherlands, July 23-27, 2007

Loomer, J. (2012), Facebook Insights Glossary of Terms [Infographic], from http://www.jonloomer.com/2012/09/11/facebook-insights-glossary/

Lovett, J. (2011). Social media metrics secrets. Indianapolis, IN: Wiley Publishing, Inc.

Lua, Alfred, 2017, How to Use Facebook Insights and Analytics to Boost Your Social Media Marketing Strategy, https://blog.bufferapp.com/facebook-insights accessed 15/08/2017

Mandelbaum, A. (2014), Sentiment Analysis as Customer Service, from: http://www.business2community.com/customer-experience/sentiment-analysis-customer-service-0760892#B3uM1vp6zDIrVUJv.97 accessed at 13/03/2017

Mäntylä, et al., The Evolution of Sentiment Analysis - A Review of Research Topics, Venues, and Top Cited Papers, 2016

Metelo, António Perez, 2015, ""Nós temos como objetivo continuar a crescer à mesma taxa de 30%", online: http:// http://iniciativas.globalmediagroup.pt/2015/11/17/nos-temos-como-objetivo-continuar-a-crescer-a-mesma-taxa-de-30/#more-1074 accessed at 10/09/2017

Moro et al., (2015), Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach

Munezero, M.D.; Montero, C.S.; Sutinen, E.; Pajunen, J. Are they different? Affect, feeling, emotion, sentiment,and opinion detection in text. IEEE Trans. Affect. Comput. 2014,5, 101–111

Musiał, K. & Kazienko, P. World Wide Web (2013) 16: 31. doi:10.1007/s11280-011-0155-z

Nadali, A; Kakhky, E.N.; Nosratabadi, H.E., “Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system” in (ICECT), 2011 3rd International Conference on Electronics Computer Technology, vol.6, no., pp.161, 165, 8-10 April 2011

Nasukawa and Yi, 2003, Sentiment analysis: Capturing favorability using natural language processing, DOI: 10.1145/945645.945658 · Source: DBLP Conference: Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP 2003), October 23-25, 2003, Sanibel Island, FL, USA

Neti, S., 2011, SOCIAL MEDIA AND ITS ROLE IN MARKETING, International Journal of Enterprise Computing and Business Systems ISSN (Online) : 2230-8849, Vol. 1 Issue 2 July 2011

Niemeier, S., Zocchi, A., and Catena, M. 2013. "Reshaping Retail: Why Technology Is Transforming the Industry and How to Win in the New Consumer Driven World." Hoboken: Wiley.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Found. Trends Inf. Retr., 2(1-2), 1-135. doi:10.1561/1500000011

Parfois, 2017, https://www.facebook.com/pg/parfois/about

Patroni J., Von Briel F., Recker J. (2016), Do retailers leverage consumer social media content for innovation? An exploratory study, Australasian Conference on Information Systems, 2016, Wollongong

Pentina, I. and Koh, A.C. (2012) ‘Exploring social media marketing strategies in SMEs’, Int. J. Internet Marketing and Advertising, Vol. 7, No. 4, pp.292–310.)

Phan, M., Thomas, R. and Heine, K. (2011), “Social media and luxury brand management: the caseof Burberry”,Journal of Global Fashion Marketing, Vol. 2 No. 4, pp. 213-222, available at:http://dx.doi.org/10.1080/20932685.2011.10593099

Pletikosa, Michahelles (2011). Understanding Social Media Marketing: A Case study on topics, categories and sentiments on a Facebook brand page, doi: 10.1145/2181037.2181066 Conference: 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11), At Tampere, Finland

Podobnik, V. (2013) “An analysis of facebook social media marketing key performance indicators: The case of premier league brands”, In Telecommunications (ConTEL), 2013 12th International Conference on, Zagreb, Croatia, 26-28 June 2013, IEEE, pp 131-138.

Polat, K. & Gunes, S. (2009). A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. Expert Systems with Applications. 36 (2). pp. 1587-1592. doi: 10.1016/j.eswa.2007.11.051.

Power D., Phillips-Wren, G. (2011) Impact of Social Media and Web 2.0 on Decision-Making, DOI: 10.3166/jds.20.249-261

Prena, Social media as marketing tool for SMEs: opportunities and challenges, Academic Journal of Business, Administration, Law and Social Sciences, Vol. 2 No. 3, 2016

Pundir, S. & Amala, R. (2015). Evaluation of Biomarker using Two Parameter Bi-exponential ROC Curve. Pakistan Journal of Statistics & Operation Research. 11 (4). pp. 481-496. doi: 10.18187/pjsor.v11i4.992.

Quinlan, J. R. (2006). Bagging, Boosting, and C4.5. Proceedings of the thirteenth national conference on Artificial intelligence. 1. pp. 725-730.

Ramona Sukhraj 2017, 27 Social Media Campaign Ideas from Big Brands You Want to Be (https://www.impactbnd.com/blog/25-social-media-campaign-ideas) accessed 25/08/2017

Ramona Sukhraj 2016, 10 Social Media KPIs You Should Track and Monitor (https://www.impactbnd.com/blog/8-social-media-kpis-you-should-track-and-monitor) accessed 04/01/2017

Ray, A. The ROI of Social Media Marketing: A Balanced Marketing Scorecard Thoroughly Validates Social Media Value, Forrester Research, July 16, 2010

Retail,I 2015."Retailers Underestimate Digital Influence." from https://www.insideretail.com.au/blog/2015/07/20/retailers-underestimating-digital-influence/

Roblek, V., Meško, M., Bertoncelj, A., Social Media, Organizational Changes and Added Value in Knowledge-Based Industries (June 7, 2013). Proceedings of the 1st AARESOC International Conference on Business & Management, June 5-7, 2013, Izmir,Turkey. Available at SSRN: https://ssrn.com/abstract=2275900

Russ Henneberry 2014, How to Measure Social Media Lead Generation by Channel, Social Media Marketing, Lead generation (https://www.marketingcloud.com/blog/social-media-lead-generation/)

Sadler, R. & R. D. Evans 2016 “Social Media Strategies in the Retail Sector: Analysis and Recommendations for Three Multi-National Retailers”

Saothawan, D., 2014, Consumer Social Media Analytics in retail-Prerequisites, business value and application for Social Media Marketing 10.13140/2.1.3174.2406.

Sarkisova G., 2013, Decoded: What Interactions On Facebook Really Mean from: http://www.complex.com/pop-culture/2013/02/decoded-what-interactions-on-facebook-really-mean/ accessed at 28/09/2017

Sayad, S. (2016). Model Evaluation - Classification. Retrieved September 16, 2016, from Saed Sayad Web site: http://www.saedsayad.com/model_evaluation_c.htm

Schivinski, B., Dąbrowski, D. (2013). The impact of brand communication on brand equity dimensions and brand purchase intention through facebook. GUT FME Working Paper Series A. Gdansk (Poland): Gdansk University of Technology, Faculty of Management and Economics, 4, pp. 1–24.

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & HerreraViedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. Information Sciences, 311(0), 18-38. doi: http://dx.doi.org/10.1016/j.ins.2015.03.040

Shamoon, Sumaira, and Saiqa Tehseen. "Brand Management: What Next?" Interdisciplinary Journal of Contemporary Research In Business 2.12 (2011): 435–441. Business Source Complete. Web. October 20, 2012.

Sigala, M., Gretzel, U. and Christou, E. (2012), Social Media In Travel, Tourism And Hospitality: Theory, Practice And Cases, VT: Ashgate, Farnham, Surrey, Burlington.

Sorescu, A., Frambach, R. T., Singh, J., Rangaswamy, A., and Bridges, C. 2011. "Innovations in Retail Business Models," Journal of Retailing (87), pp. S3-S16.

Statcounter, 2017, http://gs.statcounter.com/social-media-stats/all/portugal accessed at 15/09/2017

Statista website: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ accessed at 28/09/2017 accessed at 12/07/2017

Stieglitz, S. and Dang-Xuan, L. (2013), “Social media and political communication: a social media analytics framework”, Social Network Analysis and Mining, Vol. 3 No. 4, pp. 1277-1291.

Stelzner, M. (2013): 2013 Social Media Marketing Industry Report. How Marketers use Social Media to grow. Social Media Examiner, 2013. Available online at http://www.socialmediaexaminer.com/SocialMediaMarketingIndustryReport2013.pdf

Stephen D. Rappaport, “Listen First – Turning Social Media Conversations into Business dvantage”, John Wiley & Sons, Inc., 2011.

Sterne, J. Social Media Metrics: How to Measure and Optimize Your Marketing Investment, John Wiley & Sons, 2010

Stieglitz, S. et al., 2014. Social Media Analytics. Business & Information Systems Engineering, 6(2), pp.89–96. Available at: http://link.springer.com/10.1007/s12599-014-0315-7

Stodder, D. 2012, Customer Analytics in the Age of Social Media, https://www.tableau.com/sites/default/files/whitepapers/tdwi_bestpracticesreport_customeranalytics_july2012.pdf, pag. 25

Tan, Luke Kien-Weng; Na, Jin-Cheon; Theng, Yin-Leng; Chang, Kuiyu (2012): Phrase-Level Sentiment Polarity Classification Using Rule-Based Typed Dependencies and Additional Complex Phrases Consideration. In J. Comput. Sci. Technol. 27 (3), pp. 650–666. DOI: 10.1007/s11390-012-1251-y.

Taylor J., 2017 What is a KPI, Metric or Measure? https://www.klipfolio.com/blog/kpi-metric-measure accessed at 31/08/2017

Thelwall, Mike; Buckley, Kevan; Paltoglou, Georgios; Di Cai; Kappas, Arvid (2010): Sentiment strength detection in short informal text. In J. Am. Soc. Inf. Sci. 61 (12), pp. 2544–2558. DOI: 10.1002/asi.21416.

Thompson G., Forms and Types of Social Media, from http://socialnetworking.lovetoknow.com/Forms_and_Types_of_Social_Media accessed at 28/09/2017

Turban, E., Sharda, R., & Delen, D. (2011). Decision support and business intelligence systems: Pearson Education India.

Umar Ruhi (2014), " Social Media Analytics as a Business Intelligence Practice: Current Landscape & Future Prospects", Journal of Internet Social Networking & Virtual Communities, Vol. 2014 (2014), Article ID 920553, DOI: 10.5171/2014.920553

Varkaris, E. and Neuhofer, B. (2017) "The influence of social media on the consumers' hotel decision journey", Journal of Hospitality and Tourism Technology, Vol. 8 Issue: 1, pp.101-118, https://doi.org/10.1108/JHTT-09-2016-0058

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

Wang, X., Yu, C., Wei, Y. (2012). Social media peer communication and impacts on purchase intentions: A consumer socialization framework. Journal of Interactive Marketing, 26(4), pp. 198–208.

Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. Computer, 40(9), 96-99.

Wijnhoven, F., & Bloemen, O. (2014). External validity of sentiment mining reports: Can current methods identify demographic biases, event biases, and manipulation of reviews? Decision Support Systems,59(0), 262-273. doi:http://dx.doi.org/10.1016/j.dss.2013.12.005

Wikipedia, Emoji, https://en.wikipedia.org/wiki/Emoji acessed at 07/03/2017

Wilson, T.; Wiebe, J.; Hoffmann, P. (2009): Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. In Journal Computer Linguistics 2009 (35), pp. 399–433.

Wiltshire, Crys, 2017, The Fake Followers Epidemic, online at: https://www.gshiftlabs.com/social-media-blog/the-fake-followers-epidemic/ accessed 15/08/2017

Wu, et al. (2015). Gaining competitive intelligence from social media data. Evidence from two largest retail chains in the world.

Zailskaitė-Jakštė L., Kuvykaitė R., (2016), Conceptualizing the social media communication impact on consumer based brand equity, Trends economics and management, pag.68-74, DOI: http://dx.doi.org/10.13164/trends.2016.25.68

Zeng, D. et al (2011) Social Media Analytics and Intelligence, Article in Intelligent Systems, IEEE, DOI: 10.1109/MIS.2010.151

Zeng, D., Chen, H., Lusch, R. and Li, S.H. (2010), "Social media analytics and intelligence", IEEE Intell. Syst., Vol. 25 No. 6, pp. 13-16