الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي و البحث العلمي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
جامعة عمّار ثليجي بالأغواط
UNIVERSITE AMAR TELIDJI LAGHOUAT
كلية العلوم
FACULTE DES SCIENCES
قسم الإعلام الآلي
DEPARTEMENT D'INFORMATIQUE

***Mémoire de MASTER***

**Domaine :** Mathématiques et Informatique
**Filière :** Informatique
**Option :** Systèmes d'Information et de Décision

**Par :**

**Sarra HAMMOUTI & Khadidja SAHRAOUI**

**THEME**

# ENGAGEMENT IN SOCIAL MEDIA
# CASE OF ALGERIAN BRANDS

*Soutenu publiquement le 20/06/2019 Devant le jury composé de :*

| | | |
|---|---|---|
| Mr L. CHELLAMA | *M.A.(A)* | Président |
| Mr A. NAHAR | *M.C.(A)* | Examinateur |
| Mme H. CHERROUN | *Professeur* | Encadreur |

***N° : ........./ Année Universitaire 2018/2019***

# *Dedication*

This work is dedicated to :

- My beloved parents, who have been my source of inspiration and gave me strength, who continually provide their moral, spiritual, emotional, and financial support.
- My Precious family, my Brothers Kamel and Mohamed, my Sister Wafaa, my Sister in law Aicha , my sweets Islem, Rayen and wassim and my dear cousins Ibtihal and Imane.
- All my Teachers who gave us knowledge and shared their words of advice and encouragement.
- My partner in this work Khadidja Sahraoui, my friends, and classmates.
- Laghouat University profs and students.
- Finally, every one who contributed from near or far in this thesis. Thank you all for your support, encouragement and care.

**Sarra**

# *Dedication*

This work is dedicated to :

- The sake of Allah, my Creator and my Master, who helped me throughout my career and brought me to what I am now, my Great Teacher and Messenger, Mohammed (May Allah bless and grant him), who taught us the purpose of life.
- My precious family, to my Dad and my Mom, my two brothers Mohammed El Taher and Siradj Eddine, the reason of what I become today, thanks for your support and continuous care.
- All my teachers that I got the extraordinary chance to met them, thank you so much for every little and big thing.
- All my friends that I get to know along this beautiful journey, especially my partner in this work Sarra Hammouti.
- To every one who contributed from near or far in this thesis.

**Khadidja**

# *Acknowledgements*

Firstly, we must acknowledge our limitless thanks to Allah, the Ever-Magnificent, the Ever-Thankful for His help and bless. We are totally sure that this work would have never become truth, without His guidance.

We owe a deep debt of gratitude to our university for giving us an opportunity to complete this work.

We thank Ms.Bouzouad for accepting supervising us despite their big busyness and for their encouragements, guidelines, and kindness. Big thanks go to Mr.Chellama and Mr.Nahar for accepting to be one of the jury members and to examine this modest manuscript.

We would like to take this opportunity to say warm thanks to all our teachers and friends who have been so supportive along the way of doing our thesis.

Last but not least, deepest thanks go to all people who have had a significant impact on our graduate career by their support and encouragement. Thanks for making this thesis real.

**Sarra & Khadidja**

# مُلخّص

أصبح التسويق عبر وسائل التواصل الاجتماعي أحد القنوات الرئيسية التي تستخدمها الشركات للترويج لعلاماتها التجارية. على عكس الشركات العالمية ، تعتبر الشركات العربية مبتدئة في هذا المجال. الهدف من هذا العمل هو قياس مدى مشاركة كل من العلامات التجارية و المستهلك في تقنيات التسويق هذه على صفحات الفيسبوك للعلامات التجارية الجزائرية.

للقيام بذلك ، تعتمد الطريقة المقترحة على استخدام بعض مقاييس ارتباط المشاركة للعلامة التجارية وتحليل المحتوى الذي ينشئه المستخدمون ، وخاصة تحليل المشاعر.

تم تطبيق هذه الطريقة المقترحة على عينة لتوضيح تجاربنا على عينة مجمعة من 50 صفحة لخدمات وعلامات تجارية جزائرية ، حيث تم شرح ووسم 50,000 تعليق و حوالي 10,000 منشور من هذه العينة . تأخذ هذه القياسات في الاعتبار نوع التفاعل بالإضافة الى توزيع اللغات المستخدمة في التعليقات و قطبيتها. بخصوص منشورات العلامات التجارية ، فقد تم قياس نوع الوسائط المستعملة، نوع المحتوى ،تاريخ و توقيت النشر ، وتيرة النشر وطول المنشور. بالمقارنة مع عينة من العلامات التجارية العالمية ، أظهرت النتائج أن مشاركة العلامات التجارية الجزائرية مقبولة و مهمة من الناحية الإحصائية. بالإضافة إلى ذلك ، اقترحنا بعض التوصيات لتعزيز استخدام التسويق عبر وسائل التواصل الاجتماعي للشركات.

الكلمات المفتاحيّة :

التسويق عبر وسائل التواصل الاجتماعي ، مشاركة العلامة التجارية ، مشاركة المستخدم ، المعالجة الآلية للغة الطبيعية ، نوع الوسائط ، نوع المحتوى ، تاريخ و توقيت النشر ، وتيرة النشر ، وطول المنشور ومعدل المشاركة.

# Abstract

Social Media Marketing (SMM), which is marketing through Social Media, becomes one of the main channels used by companies to promote their brands. In contrast with world companies, Arabic ones can be considered as beginners in SMM. The aim of this work, is to measure the engagement for both brands and users/customers in such marketing technologies on Algerian Arabic Brand pages.

To do that, the proposed approach relies on using some Customer Brand Engagement (CBE) metrics and User Generated Content (UGC) analysis especially Sentiment Analysis (SA).

Our experiments are illustrated on a collected sample of 50 Algerian brand and service pages, from which 50 000 comments and almost 10 000 posts are annotated. These measurements take into account the type of the interactions in addition to the distribution of the languages used in comments and the measure of the comments polarity. Concerning the brand posts, the measurements take into consideration the media type, content type, posting day and time, frequency of posting and post length. Compared to a sample of world brands, the results show that Algerian brands engagement is acceptable and statistically significant. In addition, we have proposed some recommendations addressed to brand owners to promote the usage of such Social Media Marketing for companies.

**Keywords :** Social Media Marketing, Brand Engagement, User Engagement, Natural Language Processing (NLP), Media Type, Content Type, Posting Day and Time, Frequency of Posting, Post Length and Engagement Rate.

# Résumé

La commercialisation via les médias sociaux, qui est le marketing à travers les médias sociaux, devient l'un des principaux canaux utilisés par les entreprises pour promouvoir leurs marques. Contrairement aux entreprises mondiales, les entreprises arabes peuvent être considérées comme des débutants en Marketing via les médias sociaux. Le but de ce travail, est de mesurer l'engagement des marques et des utilisateurs/clients dans de telles technologies de marketing sur les pages du plateforme Facebook des marques algériennes.

Pour cette fin, l'approche proposée repose sur l'utilisation de certains métriques d'engagement de la marque et des clients et l'analyse du contenu généré par l'utilisateur spécialement l'analyse de sentiment .

Nos expériences sont illustrées sur la collection d'un échantillon de 50 pages algériennes de marques et de services, à partir desquelles 50 000 commentaires et près de 10 000 publications sont annotées. Ces mesures prennent en compte le type d'interaction, la répartition des langues utilisées dans les commentaires et la mesure de la polarité des commentaires. En ce qui concerne les publications liées aux produits des marques, les mesures prennent en compte le type du média, le type de contenu, le jour et l'heure de publication, la fréquence de publication et la longueur du poste. En comparaison avec un échantillon de marques mondiales, les résultats montrent que l'engagement des marques algériennes est acceptable et il est statistiquement significatif. En outre, nous avons proposé des recommandations visant à promouvoir l'utilisation de ce type de marketing par médias sociaux pour les entreprises.

**Mots-clés :** Marketing des médias sociaux, engagement de la marque, engagement de l'utilisateur, traitement du langage naturel, type de média, type de contenu, jour et l'heure de publication, fréquence de publication, longueur du publication et taux d'engagement.

# Table of Contents

# Table des figures

# Abbreviations

| | |
|---|---|
| **BE** | Brand Engagement |
| **CBE** | Customer Brand Engagement |
| **CE** | Customer Engagement |
| **CGC** | Customers Generated Content |
| **CT** | Content Type |
| **CR** | Comment Ratio |
| **ALGD** | Algerian Dialect |
| **ECT** | Entertainment Content Type |
| **ER** | Engagement Rate |
| **FP** | Frequency of Posting |
| **ICT** | Information Content Type |
| **ID** | Interaction Duration |
| **LE** | Large Enterprises |
| **LR** | Like Ratio |
| **ME** | Medium-sized Enterprises |
| **MSA** | Modern Standard Arabic |
| **MT** | Media Type |
| **NLP** | Natural Language Processing |
| **PD** | Posting Day |
| **PER** | Post Engagement Rate |
| **PT** | Posting Time |
| **RCT** | Remuneration Content Type |
| **SA** | Sentiment Analysis |
| **SE** | Small Enterprises |
| **SM** | Social Media |
| **SMA** | Social Media Analytics |
| **SME** | Social Media Engagement |
| **SMM** | Social Media Marketing |
| **SR** | Share Ratio |
| **UGC** | User Gnerated Content |
| **WOM** | Word Of Mouth |

# Introduction

Social media or as it called new media covers a fundamental role in our everyday lives, becoming since its diffusion one of the most commonly used communication and information methods for people of all ages, genders, nationalities and religious beliefs. Nowadays, Social Media (SM) are being used to transmit multiple messages and are able to reach global audiences (Bartoletti, 2013).

The presence of companies on social networks especially Facebook, is being strengthened day after day, as it is not possible for the heads and managers of companies to dispense with their presence on these digital platforms. A fierce competition has emerged between companies to win as many fans as possible on SM, create a positive image of their business or any related event and facilitate their outreach and ability to keep in touch and learn more about their customers (Bartoletti, 2013). To achieve this goal, companies do not hesitate to allocate significant budgets to support their pages by advertising and employing specialists in their development according to market requirements.

Marketing which occurs via SM is known as Social Media Marketing (SMM). It can be simply defined as the use of social media channels to promote a company and its products. SMM has made possible for companies to reach targeted consumers easily, effectively and instantly. On the one hand costs lower than other marketing platforms such as face-to-face salespeople or middlemen or distributors, and on the other hand enhancing reach to customers that may not be accessible due to temporal and locational limitations of existing distribution channels (Nadaraja and Yazdanifard, 2013).

It's important for businesses to join in and show that they are listening to their customers by participating in discussions on their own platforms and those of their followers, which make Social Media Engagement (SME) an essential part of social media marketing. It's the process by which online communications and the content that companies post online help build connections with other people within online communities, build relationships with others that ideally result in some kind of reaction, interaction, or action (Sherman and Smith, 2013). SMM considered as trigger for buying intentions and decisions and allow the creation and exchange of User Generated Content, that's why the brand owners should make the SMM a core part of their marketing strategy because conceptual studies suggest that community engagement leads to increased levels of brand engagement (Dessart, 2017).

Throughout the past decades, there have been enormous advancements in computer technology, electronics and telecommunication. In particular, advancements made in the storage, analysis and retrieval of vast amounts of data have been occurring at an exponential rate (Bartoletti, 2013). Social media data definitely belongs in the big data category, therefore, all the challenges under the definition of big data are applied to social media data as well (Gonçalves, 2017). This in turn, has led to the growth of database technology that has allowed companies to collect very useful information on customers and their buying behavior. Social media data is different from traditional documents such as newspaper articles, these non-structured texts can be found in many formats, written by different people in many languages and styles, written in everyday language. Moreover, authors are not professional writers and come from thousands of places. So it is a scientific challenge to develop powerful methods and algorithms which extract relevant information from a large volume of data in different languages.

Billions of comments and reviews are added to the web each day, which has led to the need to mine users opinion in order to discover useful information. Here where Natural Language Processing (NLP) methods in information extraction, automatic categorization and clustering, automatic summarization and machine translation need to be adapted to a new kind of data (Bartoletti, 2013). In addition, a new thematic of NLP, known as Sentiment Analysis (SA) or opinion mining (OM) is to extract users sentiment/opinion from created contents. Nowadays, SA is used mainly by businesses to discover the opinions of different customers as part of marketing purposes (Mataoui et al., 2016).
This layer of Social Analytics uses Natural Language Processing (NLP) to understand whether social conversations are positive or negative, and to measure the strength of those emotions (Bartoletti, 2013).

For the Algerian case, it might be challenging because Algerian Arabic or ALGD is considered as one of the most "hard to understand" Arabic dialects varieties, it is far less normalized and standardized compared to Modern Standard Arabic (MSA). It has a vocabulary inspired from Arabic but the original words have been altered phonologically. In addition, a rich vocabulary consisting of foreign words of French origin are an essential part of the spoken language of Algerians Phonology, morphology, lexicon and syntax of ALGD which is very difficult to understand for the citizens of the other Arab countries. The first feature of ALGD is the use of words that comes from several languages, the second feature is related to the use of Arabic expressions encoded in Romanized Arabic or foreign expressions (mostly French) encoded in Arabic letters, the third feature is the combination of the two first features, the latter feature is related to the use of words written in a very specific form, the form that most Algerians generally used for writing short messages (Mataoui et al., 2016).

So the objective of our thesis is to design an approach that is twofold. On the one hand, we aim to measure the brand engagement by analyzing the content of social networks taking the example of Facebook pages concerning the Algerian brands. This measure aims to know whether they are well engaged or not in terms of selected metrics compared to the world brands and what are the factors that influence on the engagement. On the other hand, the same thing will be explored for the customer engagement. In addition, we have investigated measuring their satisfaction toward the brand by observing their sentiment whether it's negative, positive or even neutral.

This thesis is structured as follows :

1. Chapter 1 introduces some generalities : definitions and concepts about SM, SMM, the User Gnerated Content (UGC). It also introduces the need of NLP tasks to analyze and extract information especially the SA field. Then, we introduce notion of SME either by the brand or by the customers and the used metrics.

2. In chapter 2, we review some related work in the aim to understand their methodology to measure the engagement in social media and approaches that have dealt with SA.

3. In chapter 3, we focus on the description of our Customer Brand Engagement (CBE) measurement process and the obtained results narrowed on a sample of Algerian brands collected dataset.
   We conclude and expose some future work.

# Chapter 1

# Generalities

In this chapter, we present concepts about SM concerning the marketing side on one hand, and the analytical side on the other hand. We will mention also the engagement in social media according to the brand communities and their customers, highlighting some existing approaches, metrics and factors that help for engagement measurements.

## 1 Definitions and Concepts

This section aims to give a definition of the concept of social networks, as a start we define and introduce what are SM, SMM and UGC.

### 1.1 Social Media

Social Media (SM) is the term often used to refer to new forms of media that involve interactive participation (Harvey, 2014a). It consists of websites such as forums, blogs, video sharing websites, collaborative coding websites (e.g., GitHub) and social networking sites (e.g., Twitter, Facebook, Google Plus, LinkedIN ... ) (boyd and Ellison, 2007).
These websites allow users to create profiles, share connection and interact between them (boyd and Ellison, 2007), generate and publish content which take the form of text, image, audio and video ... etc (Tsagkias, 2012).
Kaplan and Haenlein (2010) defined SM as a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of UGC.
So, the concept of SM refers to websites and applications that enable users to create and share content or to participate in social networking.

### 1.2 Social Media Marketing

Social Media Marketing (SMM) has became as a powerful way in which businesses are reaching out to targeted customers easily. It can be simply defined as the use of existing SM platforms to build a relationship and conversation between brand owners and customers, it is also about receiving and exchanging perceptions and ideas (Drury, 2008).

### 1.3 User Generated Content

According to Harvey (2014b), the term User Generated Content (UGC) refers generally to content that is produced and disseminated by unpaid non professionals, although there is no formal definition of UGC. The Organization for Economic Co-operation and Development has written that UGC is creatively produced by active web contributors, independent of journalistic practices and professional routines.
UGC is the term used to describe any form of content such as text, posts, comments, reviews, images, video, blogs, audio files and other forms of media that are created by consumers or end-users of an online system or service, and is publicly available to others consumers and end-users.

## 2 Social Media Analytics

Social Media Analytics (SMA) has become a very broad area of research, given the widespread of social networks and their impact on the society. It helps analysts, brands, agencies, and vendors to make sense of the shift in social media to focus on what is most important, it serves to generate economic value through analysis of the meaningful and relevant signals in social media data (Phillips, 2014).

SMA is the process of *Capture* (Data identification), *Understand* (Data analysis), and *Present* (Information interpretation) (Fan and Gordon, 2014) that aims to transform any social media data to meaningful and understandable messages. Figure 1.1 shows the main tasks for the SMA process. A lot of tools and techniques are developed in such purpose,

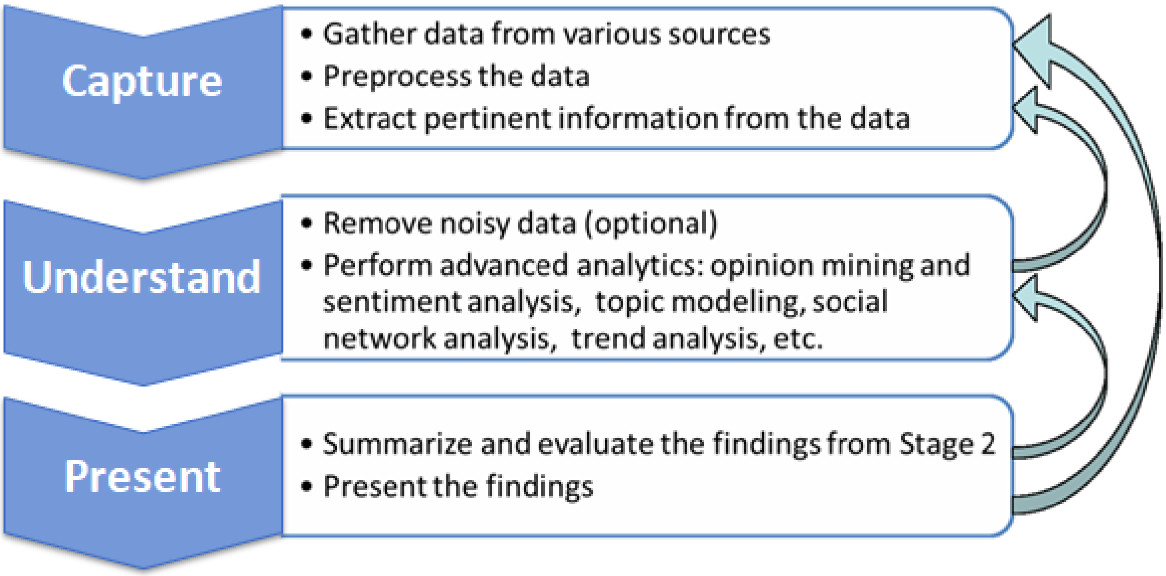



FIGURE 1.1 – Social media analytics process (Fan and Gordon, 2014)

among them we can find those that allow for the analysis of textual data such as the *natural language processing* and the *sentiment analysis*.

**Natural Language Processing**

Natural language is the language which is used or spoken by the human being, these languages are Arabic, English, French, Spanish, Korean and so on.
NLP is a field of computer science and linguistics concerned with the interactions between computers and human natural language.
Natural language generation systems convert information from computer databases into readable human language. NLP is a significant area of artificial intelligence because a computer would be considered intelligent if it can understand the commands given in natural language instead of conventional programming languages (C,Fortran or Pascal) and used for creating language translator, Machine Translation (MT) . . . (Kumar and I. K. International Publishing House, 2012).

#### 2.0.1 NLP System

NLP systems are often big software engineering projects, success requires that systems can be improved incrementally. NLP has two parts : understanding (input side processing) and another part is generation (output side processing). It needs natural language as input and gets another natural language as output. A NLP system follows some steps which are shown in below : (see Figure 1.2).

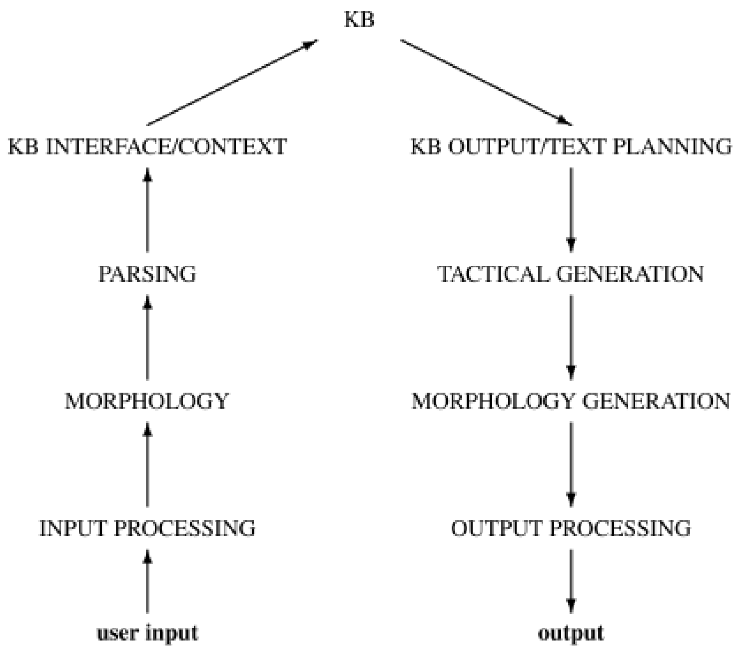



FIGURE 1.2 – NLP System (Martin, 2004)

1. Input Preprocessing : the input to the main NLP component is segmented text.
2. Morphological analysis : this is relatively well understood for the most common languages that NLP has considered, but is complicated for many languages (ex : Turkish, Basque).
3. Parsing : this includes syntax and compositional semantics, which are sometimes treated as separate components.
4. KB(Knowledge Base) Interface/Context module : this maintains information about the context, for anaphora resolution, for instance.
5. KB output/Text planning : the part of language generation that's concerned with deciding what meaning to convey.
6. Tactical generation : converts meaning representations to strings. This may use the same grammar and lexicon as the parser.
7. Morphological generation : as with morphological analysis, this is relatively straightforward for English.
8. Output processing : text-to-speech, text formatter, etc. As with input processing, this may be complex, but for now we'll assume that we're outputting simple text.

#### 2.0.2 Major tasks in NLP

There are many NLP tasks, Jurafsky and Martin (2009) presented the major ones which are :

- Spelling and Grammar Checking : Checking spelling and grammar suggesting alternatives for the errors.
- Summarization : Generating a short summary from one or more documents, sometimes based on a given query.
- Machine Translation : Automatically translate a document from one language to another.

- Question-Answering : This is a generalization of simple Web search, where instead of just typing keywords, a user might ask complete questions.
- Speech Recognition : Recognizing a spoken language and transforming it into a text.
- Word Prediction : Predicting the next word that is highly probable to be typed by the user.
- Information Extraction : Like sentiment analysis, polarity granularity and opinion target module.
  - Sentiment Analysis :The extraction of subjective information from a data source using natural language processing, computational linguistics and text analytic systems (Batrinca and Treleaven, 2015).
  - Opinion Mining : Making automatic systems able to determine human opinion from text written in natural language (Batrinca and Treleaven, 2015).

## 3 Social Media Engagement

Social media engagement (SME) refers to user interactions such as liking or commenting on any social media content that shows their interest in that content.

In marketing context, Dessart (2017) has defined the term as the state that reflects the positive reactions of consumers towards the community and the focal brand. These reactions are expressed through varying levels of affective, cognitive and behavioral manifestations.

### 3.1 Customer Brand Engagement

Customer Brand Engagement (CBE) becomes a new perspective in managing relationships with consumers, it gives an opportunity for companies to explore the values of their customers. Studies on consumer engagement have provided various definitions of the concept, for example Hollebeek (2011) has defined it as the level of a customer's cognitive, emotional and behavioral investment in specific brand interactions and its also defined by Leckie et al. (2016) as customer involvement, participation and self-expression toward a brand.

### 3.2 Brand Communities

A brand community is an important platform for consumer engagement. It is a group that consists of individuals who are specialized, non-geographically bound community based on a structured set of social relationships among admirers of a brand (Muniz and O'Guinn, 2001). Brand communities facilitate interactions through exchange of opinions about the brand or a particular product among consumers (McAlexander et al., 2002).

### 3.3 Approaches for Measuring Engagement in Social Media

Engagement in social media, is a multifaceted complex phenomenon that can be measured by a number of potential approaches (Lalmas et al., 2014) such us :

- Self-Reporting Approaches : which consist of asking people about their experiences.
- Physiological Approaches : Which assess how the body functions, the important ones are those related to cognitive or affective states that can be captured by sensors, cameras, or software such as *Eye tracking*, *Mouse pressure* and *Biosensors* etc.
- Web Analytics Approaches : Which refer to the extraction of parameters thought to affect user engagement, from the digital traces left by users while interacting with a website.

Each approach includes several techniques for measurement. Figure 1.3 illustrate the most important ones for each approach.

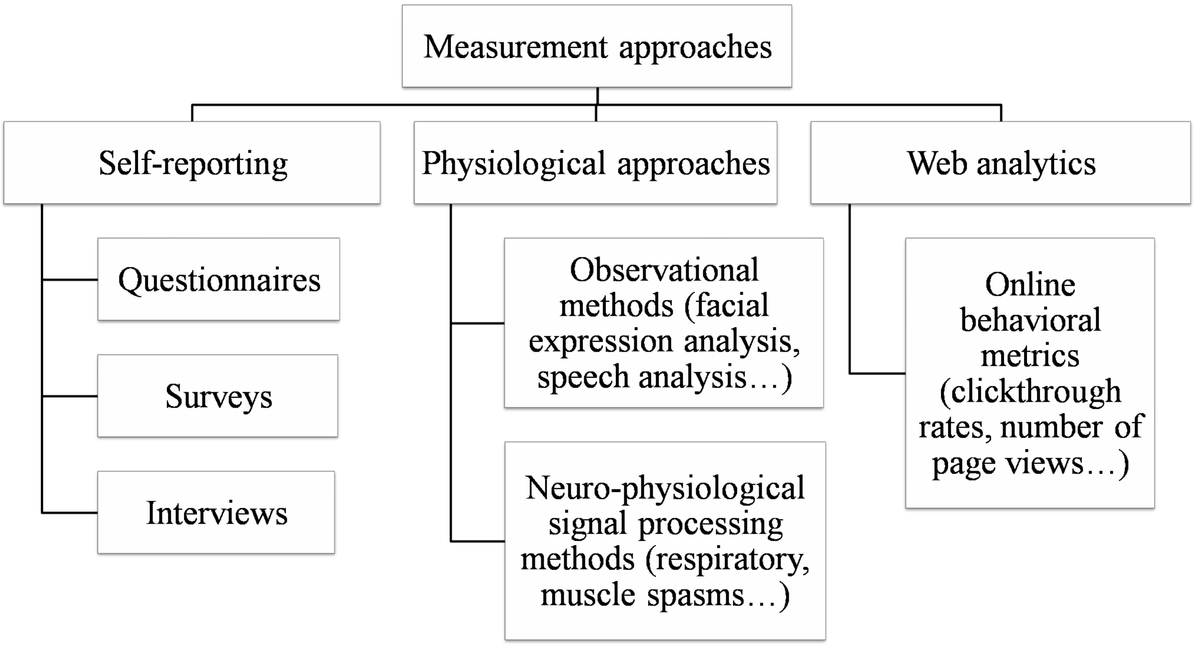



FIGURE 1.3 – Measurement approaches (Lalmas et al., 2014).

For our purpose, we have opted to deal with the web analytics approach and for that we have to identify all the related engagement metrics that can be used and measured.

### 3.4 Social Media Engagement Metrics

SMEcan be measured using several metrics depending on the targeted SM platform. According to Perreault and Mosconi's (2018) literature review on SME, the actions that may measure CBE are :

- Number of reactions : including likes, loves, hahas, wows, sads, angrys in Facebook as illustration.
- Number of comments and shares.
- Number of followers : Number of users who follow a page.
- Number of fun : Number of users who like a page.
- Number of views : including page views which is the number of times a Page's profile has been viewed by people or the number of people who had viewed any content from a Page or about it enters their screen. This is called the *Reach*, it can be also the number of times any content from a Page or about it entered a person's screen which is named the *Impression.*
- Click-through-rate : The number of times a post or a page was clicked.
- Posting on the page : Number of post on the page in a period.
- Conversion or Buy : The number of times a specific action like purchase or add to cart was taken

In addition to these straight metrics, there is a fundamental one called *"Post Engagement Rate"*. According to Facebook this metric has different ways of measuring, here's are two of them :

$$PostEngagementRate = \frac{reactions + comments + shares}{number of followers} * 100$$

$$PostEngagementRate = \frac{reactions + comments + shares}{Reach} * 100$$

Where $reactions$ : is the total number of post reactions, $comments$ : is the total number of post comments, $shares$ : is the total number of post shares, $followers$ : is the total number of followers on the day of posting.

Although, the second formula gives a more relevant result than the first, but it uses the *reach* metric which is considered as a private data (visible only to the platform and the page owners or some measurement organizations that have taken the permission of access), thus it can't be applicable by simple users.

Several research have made in this area, they have proved that these metrics are not absolute but they are related to other influencing factors. Among these factors we can found those presented by Barger et al. (2016) and Perreault and Mosconi (2018) in their marketing literature reviews and by Pletikosa Cvijikj and Michahelles (2013), Here are some of them :

- **Brand Factors** : including Trust the brand popularity, the quality of its relationship with its customers and the degree of their trust on it etc.
- **User Factors** : including their culture, attitude toward social networking sites, Age, gender, etc.
- **Technical Factors** : including time frame (posting time), day of the week (posting day), length of the wall post, Frequency of Post, etc.
- **Page/Content Factors** : Include customers Attitude toward content, emotional sentiment of message, content type (informative, entertainment, etc.), content format (textual, photo, video, etc.), page usefulness... etc.

## 4 Conclusion

In this chapter, the basic concepts concerning SM, SMA, SME are presented and explained. These concepts, helps us to get a deep vision on our study topic and the possible challenges we face, which allow us to address the next chapters.

# Chapter 2

# Related work

In this chapter, we will present some literature work that are reviewed and which are related to the social media engagement measurement and/or to the sentiment analysis field.

# 1 Engagement in Social Media

Many researchers have investigated the study of CBE in SM. They aimed measuring the related engagement by processing some new metrics. In general, these studies focused on Facebook pages and personality related content. There are various work and studies that have done in this field, like Karjaluoto et al. (2015), Dessart (2017) and Quesenberry and Coolsen (2018). We have chosen to review these ones : Pletikosa Cvijikj and Michahelles (2013), Jayasingh and Venkatesh(2015) and Olczak and Sobczyk(2013). Our review focused on highlighting the used metrics, the defined/studied factors and the nature of SM data.

## 1.1 Pletikosa Cvijikj and Michahelles (2013)

The work of Pletikosa and Michahelles has focused on analyzing influencing factors in terms of content characteristics created by companies. These factors include MT, CT, PD and PT over the level of online customer engagement measured by the quantities of likes, comments, shares and Interaction Duration (ID) with the Facebook brand page.
The authors developed a model which explains the relationship between these constructs. Figure 2.1 illustrates it.

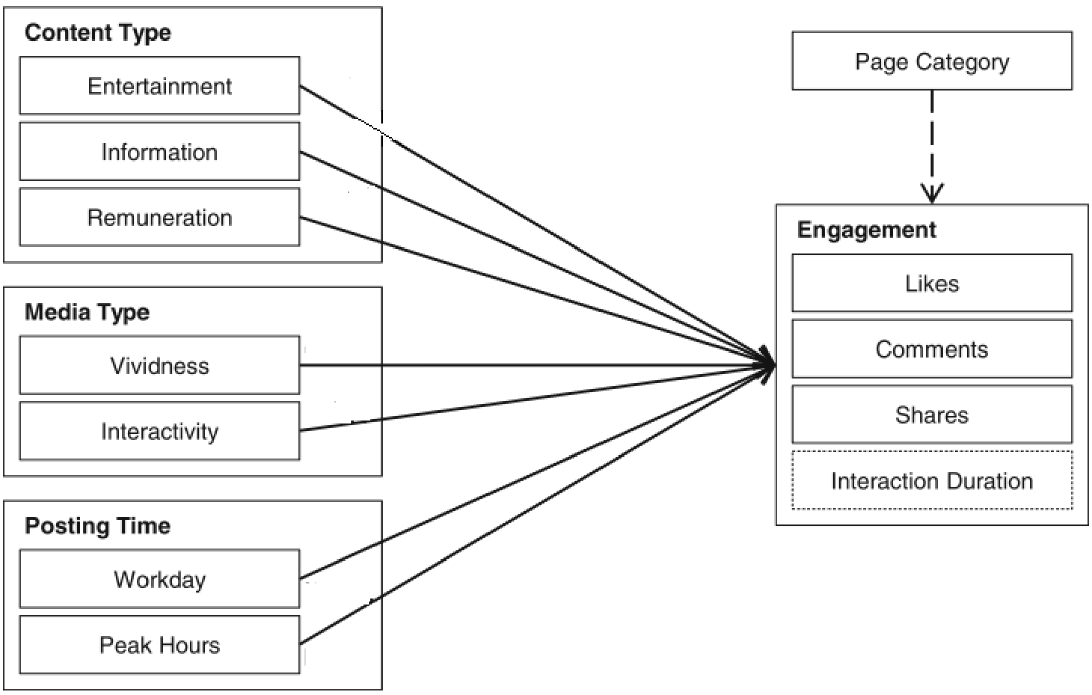



FIGURE 2.1 – Conceptual framework for relations between post characteristics and online engagement (Pletikosa Cvijikj and Michahelles, 2013).

Their Experimental study is done on 100 sponsored brand pages. Their obtained results showed that :

1. Providing entertaining and informative content significantly increases the level of engagement.

2. In addition, fans positively react to content offering remuneration but only in commenting form.
3. They also showed that vividness, which refers to the way a post could appeal to individuals' senses, increases. While interactivity, which is defined as the extent to which communication parties could act on one another on the communication medium and on the messages interchanged, decreases the level of engagement over moderator posts.
4. Photos is the most appealing post media type.
5. Finally, posts created on workdays increase the level of engagement, while posting in peak activity hours will reduce the level of engagement.

### 1.2 Olczak and Sobczyk (2013)

Work of Olczak and Sobczyk has analyzed the Brand Engagement (BE) and fans engagement on Facebook within four European countries ( Poland, France, Germany and the United Kingdom) based on four mobile phone operator's activities (Orange, T-mobile, Virgin and o2). They have introduced the strategy of posting in order to achieve the best results in getting engagement. The strategy combined of several elements :

- **Posting time** : Their statistics proved that the posts on weekends (respectively outside of business hours) perform about 16% (respectively 20%) better than workdays (respectively business hours).
- **Frequency of posting** : They have added that optimal post frequency is a separate question depending on a Page's audience, content production skills and post lifetime. However and based on other studies, they assumed that Facebook Pages shouldn't post more than one time every three hours.
- **Length of post** : They deduced that the best practice for text posts's length is between 100 and 250 characters (less than 3 lines of text), which generates about 60% more likes, comments and shares than posts greater than 250 characters.
- **Content type** : They considered 3 types of content :
  1. Related to products/services : such as information about the new offer.
  2. Related to the brand : containing photos from new store launch or contests with a chance of winning brand's souvenirs.
  3. Unrelated to the brand : usually connected with seasonal events, like holidays, sports games, entertainment, etc.

  Figure 2.2 presents strategies for posting on Page regarding posts' content in a response to previously mentioned goals and motivations of Facebook users.
- **Promoting brand and products** : Where directly inform about the offer.
- **Express through photos** : Might be the picture of a new smartphone available in business tariffs or for example a photo that **humanize the brand**.
- **Post funny things** : Knowing that entertainment is the strong motive to join the community, there is a need to publish funny posts, like cartoons or some thing **Topical**, which might be connected with holidays like the suggestion to send a text message on Valentine's day or buy a mobile phone as a gift for Christmas.
- **Sharing validation** : Such as success stories, achievements, posts about the award, the brand has won or information about celebrities who decide to use the network's service.
- **Educate the fans** : Like explaining them the features of smartphones.

For the fans, they developed conclusive rates to determine the relation between


FIGURE 2.2 – Strategies to increase brand engagement on Facebook (Olczak and Sobczyk, 2013).

posts, likes/shares and the amount of Page's fans. These rates are defined by the following formulas :

$$PERL = (\sum(\frac{l}{p}))/Ft$$

$$PERS = (\sum(\frac{s}{p}))/Ft$$

Where $PERL$ : Post Engagement Rate for "likes", $l$ : the sum of likes at a specified period of time, $PERS$ : Post Engagement Rate for "shares", $s$ : the sum of shares at a specified period of time, $p$ :the sum of posts at a specified period of time, $Ft$ : the average number of fans at a specified period of time.
Comments (the third post's engagement feature) were not taken into consideration due to the fact that one fan can write unlimited number of comments, they could be both positive or negative and what is the most important they can be posted by Page's administrator as well as a reply to users comments.

In addition, they measured and discussed the interaction of page owners with fans, evaluating this interaction by asking the same question on Pages of mobile network operators with the four languages. The average waiting time was about an hour, the slowest turned to be almost 5.5 hours.
As a final results, they have found that leaders in Post engagement Rate failed the interaction test.

### 1.3 Jayasingh and Venkatesh (2015)

The work of Jayasingh and Venkatesh has identified the factors influencing the Customer Engagement (CE) in Facebook brand pages. They have used the Fanpage Karma [1] as SM monitoring tool to collect the data and provided a conceptual framework that helps to understand the factors influencing the CE in Facebook brand pages. Figure 2.3 illustrates their conceptual framework of CBE where a sample of 134 Indian Facebook brand pages were monitored regarding the brand's activity (posts : Status, video, picture, link, offer) as well as the customers' interactions with the brand's activity (likes, comments, etc.).

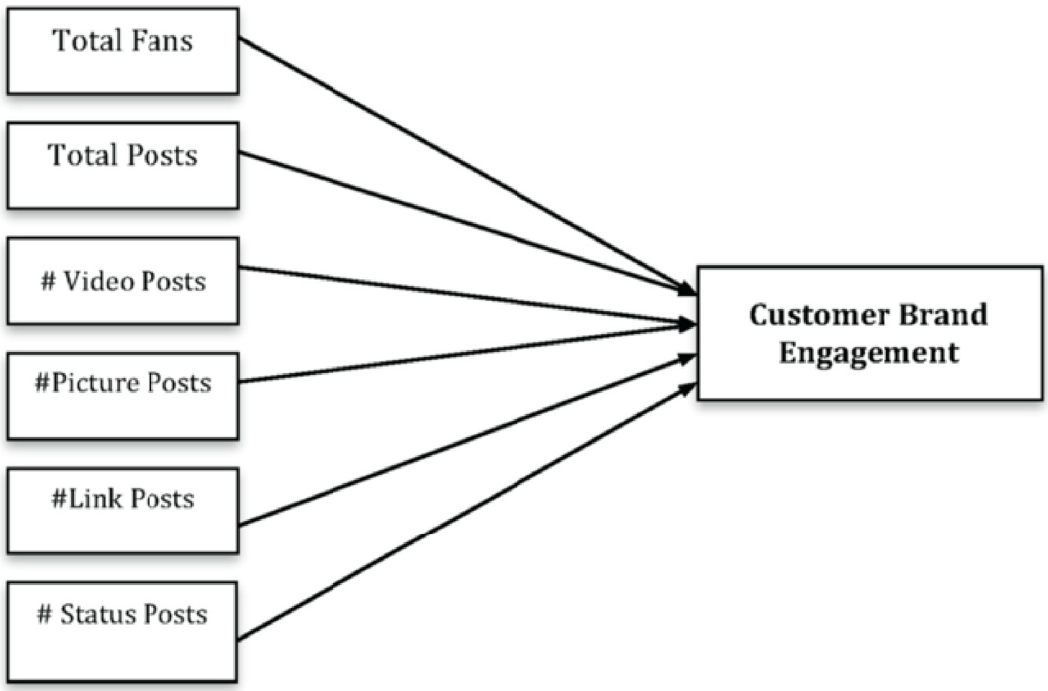



FIGURE 2.3 – Conceptual framework of customer engagement (Jayasingh and Venkatesh, 2015).

The results show that :

1. Average interaction is highest for offers and least for links.
2. Photos and offers increase the average customer interaction.
3. The determinant keys for customer engagement are content-related and frequency of brand posting activities. By this result, they prove that the number of fans following the page influence less on customer engagement as it can be.

Table 2.1 summaries these reviewed work on engagement. This summary prove that influencing factors on engagement are : Content type, Posting time, Media type, Frequency of posting, post length.

1. Fanpage Karma is an online tool for social media analytics and monitoring. It helps social media managers and agencies to engage fans better and reach a growing community in the social networks. Fanpage Karma provides valuable insights on posting strategies and performance of social media profiles like Facebook, Twitter or YouTube.

| Work | Dataset | Platform | Metrics & Factors | Main findings |
|---|---|---|---|---|
| Pletikosa Cvijikj and Michahelles(2013) | 100 Brand pages collected in period of 2 Months | Facebook | Content type, Media type, posting day and time | Entertaining and informative content increases the level of engagement.<br>Posts created on workdays increase the level of engagement.<br>Photo is the most appealing post media type.<br>Posts created on peak activity hours reduce the level of engagement |
| Olczak and Sobczyk(2013) | 10 pages belongs to 4 mobile brands collected in period of 21 Days | Facebook | Number of likes, number of shares and posting time. | Leaders in Post Engagement Rate Post Engagement Rate (PER) failed in the interaction test. |
| Jayasingh and Venkatesh(2015) | 10169 Posts of 134 Brand pages collected in period of 12 Months | Facebook | Number of fans, Customer interaction and Posts type | Higher interaction in posts containing offers, and less one in posts containing Links.<br>Content related and frequency of posting are the determinant keys for CE.<br>Photos increase the customer interaction.<br>Number of fans influence less on CE. |

TABLE 2.1 – CBE Related Work

## 2 Sentiment Analysis

The content analysis is proven that leverages understanding the user/brand behavior on SM in general and SMM specially. For that, we have reviewed some work that performed sentiment analysis and resource building.
In addition and for the purpose of our study, we have chosen to review work that dealt with Arabic Algerian SMM : **Mataoui et al. (2016)**, **Soumeur et al. (2018)**, **Rahab et al. (2017)** and **Guellil et al. (2018)**

### 2.1 Mataoui et al. (2016)

Mataoui et al. have proposed a new lexicon-based approach for vernacular Algerian Arabic sentiment analysis. This approach attempts to address the specific aspects of this very particular Arabic dialect. They mentioned the main four issues related to these features and proposed a process to handle each of these aspects :

- They proposed an approach composed of four modules : common phrases similarity computation module, pre-processing module, language detection and stemming module, and polarity computation module.
- They built a lexicon which is composed of three parts : keywords lexicon (L1), Negation words lexicon (L2), intensification words lexicon (L3). These three lexicons are enriched by a dictionary of emoticons and another dictionary of common phrases.
- Concerning the third point, they implemented a parser based on the three following steps : tokenization, normalization and stop-words removal.
- They collected posts and comments of the well-known and frequented facebook pages, this dataset for experimental purposes was filtered and annotated by experienced users in order to facilitate the evaluation process of their proposal.

In total, they have selected 206 posts and their 7698 related comments.
Experimental results show that their system obtains good performance with 79.13%.

### 2.2 Rahab et al. (2017)

The work of Rahab et al. has presented ARAbic Algerian Corpus for Opinion Mining (ARAACOM), for sentiment analysis of Algerian Arabic Newspaper and for identification of opinionated sentences. Comments have been collected from the Algerian daily Echorouk and Elkhabar. The proposed approach acts in four general steps :

1. Corpora Manual Preprocessing
2. Main Article Topics identification : for each article they take the title and try to find it one or two main topics, and recognize the substitutions of each topic that can be used by reviewers to express opinions.
3. Comments with Topic Oriented Sentences Extraction : divide the sentence into 4 elements : **Predicate, Source, Subject, Content.** They kept from comment only opinion passages which contain the important amount of sentiment orientation information (Content).
4. Opinion words identification : a sentimental lexicon is used as input to this step, this lexicon is manual prepared from positive and negative words.

**Evaluation Results :** the results are improved using the Bi-gram model, also the Binary Term Occurrence (BTO) model achieve the best results of 100%.

## 2.3 Soumeur et al. (2018)

In their framework, Soumeur et al. have tackled the problem of SA of Algerian Facebookers comments on published pages belonging to various companies.

They studied the specificity of ALGD and the linguistic behavior of Algerians on SM. They have selected more than 20 brand pages of companies in Algeria. For each page, they downloaded 250 publications, for each of these they downloaded 350 comments so at the end they obtained more than 100,000 comments using the Facebook graph API.
Then, they built a corpus of more than 25000 manually sentiment-annotated comments as positive, neutral or negative. This corpus went through a pre-processing phase (see Figure

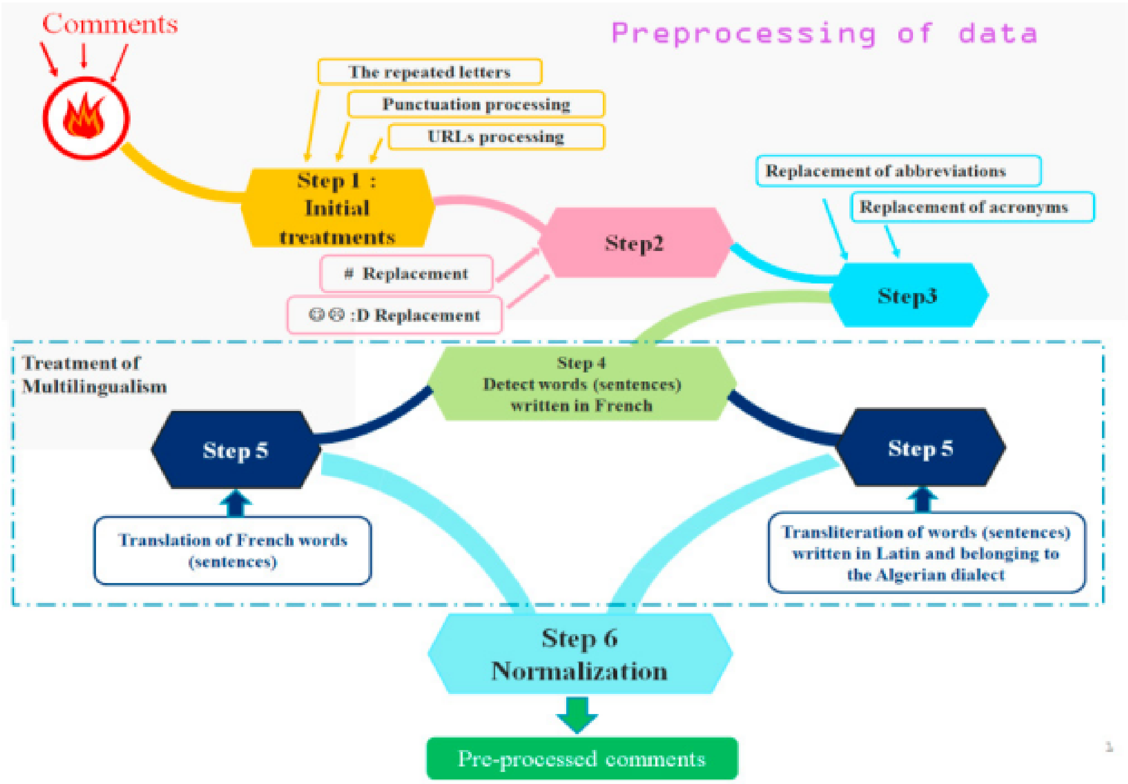



FIGURE 2.4 – Corpus preprocessing steps (Soumeur et al., 2018).

2.4), each step was evaluated in terms of its impact on the quality of the data using a Naïve Bayes classifier. They noticed the strong impact of this phase on the results they obtained, the performance significantly increasing after each step. They have implemented two types of deep neural networks, the first is a (deep) MultiLayer Perceptron (MLP) where the best configuration gave an accuracy of 81.6% and the second is a Convolutional Neural Network (CNN) that reached an accuracy of 89.5%.

## 2.4 Guellil et al. (2018)

In their work, Guellil et al. have presented an Automated Corpus Annotation for Algerian Sentiment Analysis of ALGD messages, they used Socialbakers website to collect the name of the 226 most famous Algerian Facebook pages including Ooredoo, HamoudBoualem, Algeria Telecom and Ruiba collecting 15,407,910 messages where 7,926,504 are in Arabic and 3,976,700 are in Arabizi where only taking 8000 messages. The proposed approach acts in three general steps (see Figure 2.5) :

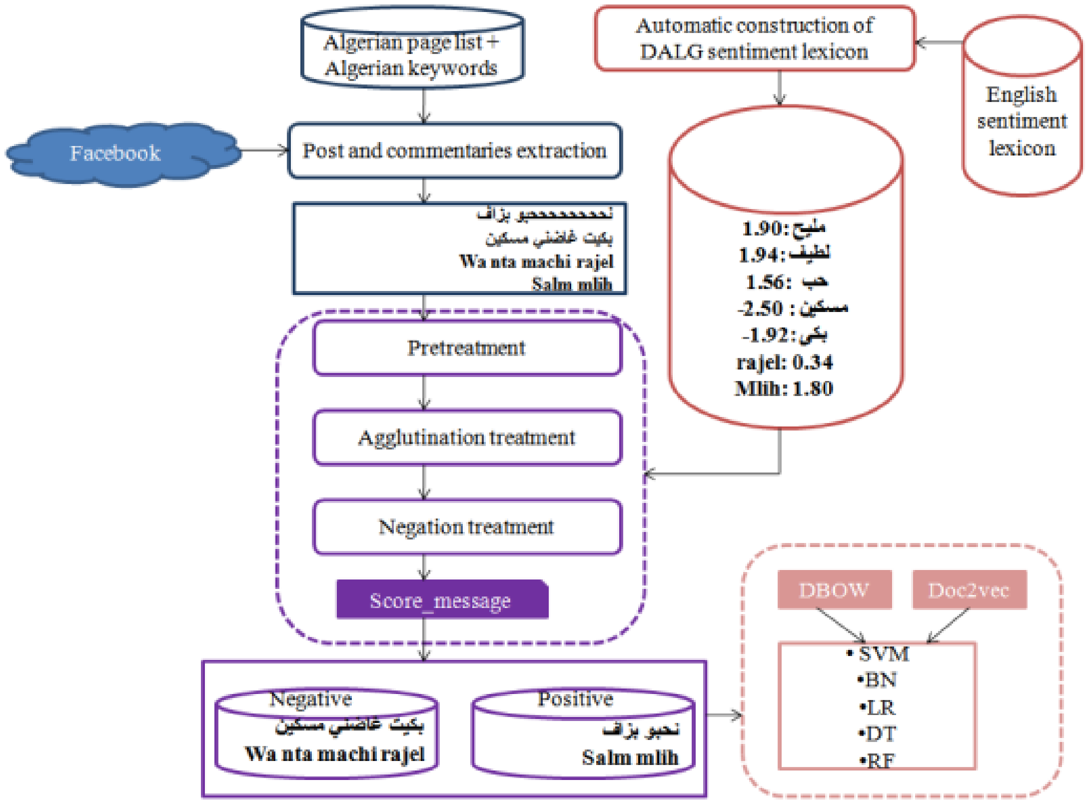



FIGURE 2.5 – The general architecture of sentiment analysis ALGD approach (Guellil et al., 2018).

1. Automatic construction of ALGD sentiment lexicon : as an input : lexicon of sentiments in English (SOCAL) where each word is scored from 1 to 5, the same score is assigned to all ALGD collected words.
2. Polarity calculation of ALGD messages : automatically annotate a set of Facebook messages (extracted from Algerian pages) as positive (+1,+5) and negative (-1,-5) :
   (a) Pretreatment of the messages.
   (b) Agglutination treatment.
   (c) Negation treatment.
3. Sentiment classification of ALGD messages : using different classifiers.

**Experimental results :** Experimental results show that the system obtains good performance with 78% using the Bag of Words (BOW) vectorization and the Logistic Regression (LR) classifier.

Table 2.2 summaries the above reviewed work.

| Paper | Dataset source | Dataset size | Language | Approach | SA level | Availability |
|---|---|---|---|---|---|---|
| Mataoui et al. (2016) | Facebook | 7698 comments | Algerian dialect | Lexicon-based | Word level | Available (dataset + application code source + lexicons) |
| Rahab et al.(2017) | Algerian daily Echorouk and ElKhabar. | - | Arabic Standard, Algerian dialect | Corpus-based | Sentence and word level | - |
| Soumeur et al. (2018) | Facebook | 100K comments | Algerian dialect | Corpus-based | Word level | - |
| Guellil et al.(2018) | Facebook | +15M comments only taking 8000 messages | Algerian dialect | Corpus-based | Word level | - |

TABLE 2.2 – SA researches.

## 3 Sentiment Analysis and Engagement in Social Media

This following reviewed work deals with sentiment analysis and Engagement measurement :

### 3.1 Saragih and Girsang (2017)

Work of Saragih has proposed investigating the CE by mining the comments of fan page Facebook and tweets of Twitter in three transports online companies in Indonesia : Gojek, Grab, and Uber using an API service. The data is taken in the period between February-March 2017 where the number of treated comments are limited only to 200 comments for each source.
There were some steps that have been followed in this method :

1. Preprocessing data : this step is performed for preparing the data for analyzing the sentiment. The process includes case folding, convert emoticon, stemming, tokenization, convert negation and stop word removal.
2. Categorizing the sentiment : each comment is classified into three sentiments : positive, negative and normal. Each word in the comments is calculated based on TF-IDF (Text frequency/inverse document frequency). If the summarization of score TF-IDF positive is higher than summarization of score TF-IDF negative, it will be classified as positive sentiment, and vice versa. The comments are considered as neutral sentiment if the score positive and negative are balanced.
3. Classifying the categorize : each comment is classified into 6 categories : Quality of service, The vacancy of driver, Socializing of service, Feedback on system by driver, Feedback on system by user and other.
   The classification of each comment is performed as follows; for each word they counted it frequency in all documents. These words are then ranked. The first 300 ranked are chosen as the word list of categorized. Then each of the 300 words are mapped into six categorized as aforementioned. Likewise categorizing sentiment, this classification uses TF-IDF. The weight of categorizing is based on the words of each categorizing which is got from 300 words. The number of data which is categorized is 400 per company, 200 Facebook and 200 Twitter for Gojek, Grab and Uber.

For the sentiment analysis part, the results show that all of transport online (Gojek, Grab, Uber) have more negative sentiments than the positive ones. Where it conduct that customer tends to use social media (Facebook and Twitter tend to be similar) for complaining than saying a positive thing. The neutral sentiments also have many comments, but it is less than negative sentiment.
For the categorization, the obtained results show that the most comments is the vacancy and expansion comments, complaining from driver for company (feedback system by driver) is the second highest.
Finally, for the categorized and Sentiment part, the highest categorization is the "vacancy and expansion", the sentiment tends to be neutral. It means that the comments are just asking about the job for driver and also asking when the service is opened in their cities.
Moreover, many of the comments in this categorization who classified as neutral sentiment, yet it looks like they have positive sentiment because most of them just asking about the job for driver and also asking when the service is opened in their cities.

## 4 Conclusion

In this chapter, we have reviewed some related work that have been done concerning the engagement in social media. We conclude that the measurement of engagement depends on multiple and various metrics depending on the (SM) platform, the research domain and the available dataset.
Those conducted metrics are : Content type, Media type, Posting Day and Time, Number of fans, Customer interaction, Posts type, Number of likes and Number of shares.

In addition, we reviewed some work related to the sentiment analysis field, especially those investigating the Algerian community comments. Each of the studied approaches presented a conceptual process containing many phases to achieve the extraction of the user sentiment.

## Chapter 3

# CBE Measurement and Analysis Process

In this chapter, we describe the process that we have designed to measure the CBE on the Algerian social media communities. The following section 1 is dedicated to the whole overview of this process where we justified the made choices. Section 2 the experiments and results are reported and commented.

## 1 Process Overview

In order to get information about CBE, we have designed a process that acts the situation in two steps. Figure 3.1 illustrates the main tasks of each one. In what follow, we will

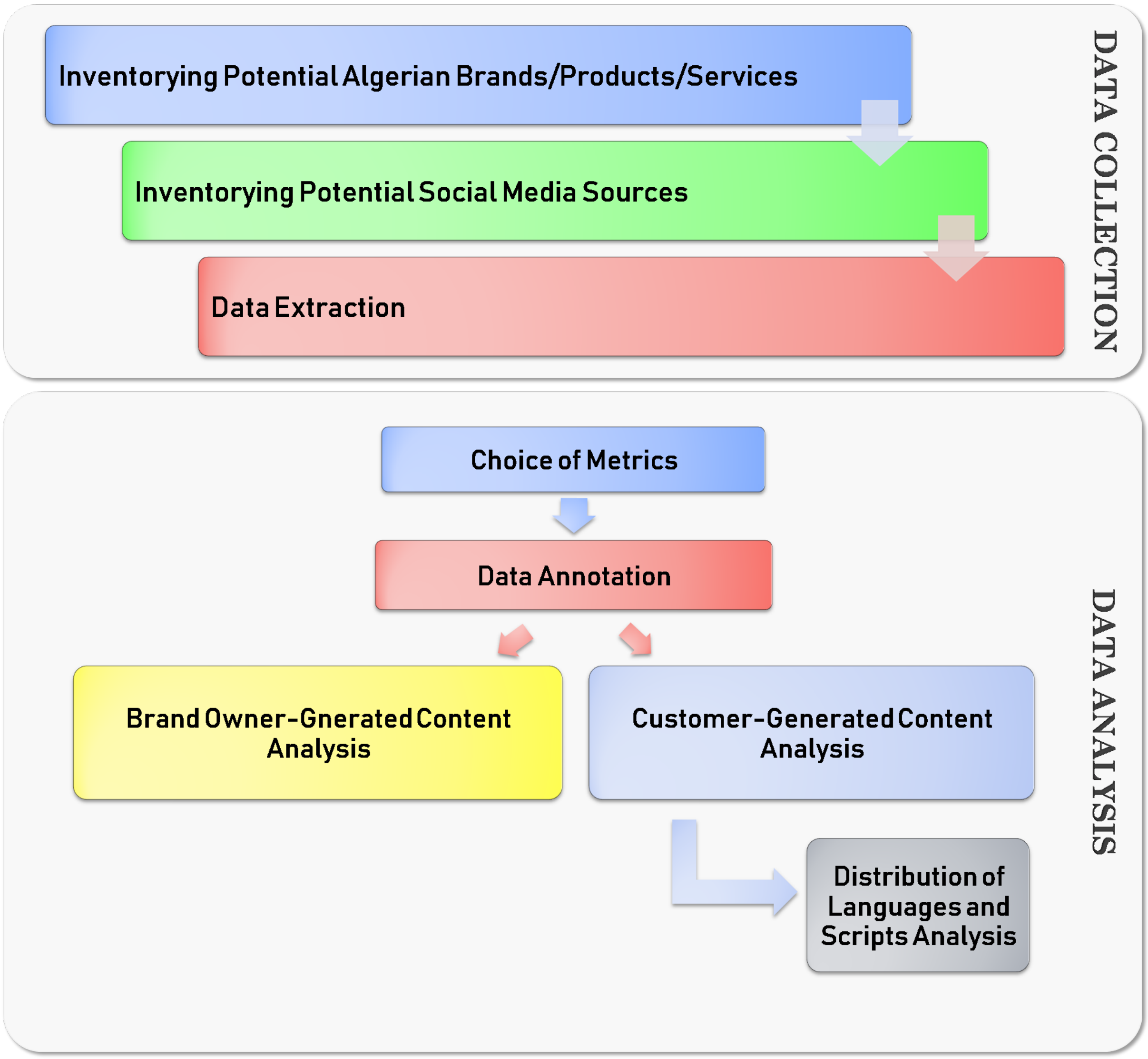



FIGURE 3.1 – CBE Process Overview

explain these steps in a formal (general) way that it can be explained for any dataset. In the experimental section, we will narrow it on our targeted dataset.

## 1.1 Data Collection

This step aims to collect a representative testbed for our system. In order to choose our sample data we have proceeded these stages :

### 1.1.1 Inventorying Potential Brands/Products/Services

First, we have to identify all brand subcategories and main services for the targeted community. In fact, we have been in touch with some economic experts for that purpose. In addition, we have chosen various brands/services with different sizes as Small Enterprises (SE), Medium-sized Enterprises (ME) and Large Enterprises (LE). This choice is guided by details and statistics that we have extracted thanks to Kompass [1] and socialbackers [2] platforms.

### 1.1.2 Inventorying Potential Social Media Sources

In this step and according to the targeted community, we have to identify the most used social media platform. We notice that some communities prefer to deal with some Web media than others depending on their culture and preferences.

### 1.1.3 Data Extraction

In order to avoid crawling useless data, this step is achieved by two stages :

1. Providing Lists : For each targeted Social Media, we define the main keywords that can help automatically searching lists. When such lists are established, a first cleaning is performed keeping only the potential suitable data. It helps to enlarge our Brand List by Brands that are well visible via Social networks (i.e. have good ranking) but not considered by experts as a powerful Brand/Service.
2. Downloading Data : in this step, we use customized scripts, based mainly on Facebook Graph API [3].

## 1.2 Data Analysis

In order to analyze our collected data, we have chosen some metrics among those that considered as efficient CBE measures. Since we didn't have enough data concerning some of these metrics, we have proceeded with an annotation phase, which took us a lot of time and effort.

Once all the required data has been obtained, we have analyzed the dataset from two levels according to the selected metrics : *brand owners generated content analysis* and *customers generated content analysis*. In addition, we have tried to extract all possible informations related to the customer behaviors on SM. Thus, we have analyzed the distribution of the used languages and the written scripts for the targeted community.

### 1.2.1 Choice of Metrics

Concerning metrics and factors mentioned in the Chapter 1, we choose the ones that can be used and measured for our need and situation. The selection based on what we have as dataset and on the related work presented on chapter 2. In fact, we have divided those metrics into two categories which are :

1. dz.kompass.com : the leading online global business directory.
2. www.socialbakers.com : social media analytics platform.
3. See : https ://developers.facebook.com/tools/explorer

1. Brand Engagement Metrics : all these metrics are related to the brand posts.
   - Content Type : including all the actual sharing action undertaken by the page moderator within a Facebook page which are :
     - Entertainement : Content gives amusement, pleasure, enjoyment or relaxation to an audience.
     - Remuneration : Content gives information about rewards, reductions, in Algerian dialect such as Tambola . . . etc.
     - Information : Content gives brand/service related informations.
     - Other : Content that dealt with condolence messages, changing page cover, . . . .

     This metric is chosen as it is important for the brand to post content that attracts consumers in order to get more engagement.
   - Media Type : Indicates the type of the post which can include Event, Link, Photo, Status and Video. It is chosen cause different media types provide different levels of interactivity so it's important to use the right media that can influence customers better.
   - Posting Day and Time : In the Facebook platform, post timing is an important aspect of scheduling.
   - Frequency of Posting : The optimal frequency, can keep customers remembering the brand without being a nuisance task for them, which conduct to more participation
   - Post Length : As a brand posts on Facebook, means it's competing with thousands of other brands for the same audience's attention. Thus, the amount of interest the audience gives to a brand is limited. So, posts that have an ideal character length have a greater chance of being read by customers and thus can provide more engagement.
2. Customer Engagement Metrics :
   - Like Ratio : Number of likes according to the total number of Fans on the day of posting, it can be calculated by the following formula :

     $$\frac{Nl}{Nf}$$

     ($Nl$ : Number of likes, $Nf$ : The total number of fans).
   - Comment Ratio : Number of comments according to the total number of fans on the day of posting, it can be calculated by the following formula :

     $$\frac{Nc}{Nf}$$

     ($Nc$ : Number of comments).
   - Share Ratio : Number of shares according to the total number of fans on the day of posting, it can be calculated by the following formula :

     $$\frac{Ns}{Nf}$$

     ($Ns$ : Number of shares).
   - Interaction Duration : To calculate the interaction duration, we used the following formula :

     $$ID = Tli - Tc$$

     (where $Tc$ is the time of post creation and $Tli$ is the time of last interaction with the post).

- Customer Sentiments : It is important for the brand community to know whether its customers are satisfied with what it offers or not. It is added that a great customer satisfaction can increase the engagement which is a sign of successful marketing strategies.

#### 1.2.2 Annotation process

According to the requirements of our proposed process mentioned on Figure 3.1 and the selected metrics, we have noticed that our collected dataset can not provide directly data for the chosen metrics such *Content Type*, *Customer Sentiments*, *Interaction Duration*, *Like Ratio*, *Comment Ratio* and *Share Ratio* and same for the *Distribution of Languages and Scripts Analysis*. For that purpose, we have proceeded with this phase to annotate both Posts and Comments to get all the required data. In fact, some of the made annotations are manually and others are automatically.

- Posts Annotation : Concerning *Content Type*, we have manually annotated all the extracted posts by giving a suitable type for each one. While, for the *Interaction Duration* we have made it using some Visual Basic macros thanks to Excel. Concerning *Like Ratio*,*Comment Ratio* and *Share Ratio*, we have calculated them using simple formula on under Excel Sheet.
- Comments Annotation : Concerning *Customer Sentiments* and the *Distribution of Languages*, we have removed all comments containing just photos and stickers because we have mainly interested to textual content and emoticons. Then, we have annotated manually a large set of them. In fact, the annotated set includes comments that belongs to different pages. For the *Used Script Analysis* we have made it using some Visual Basic macros for Excel.

#### 1.2.3 Brand Owner-Generated Content Analysis

At this level, we have analyzed the distribution of Brand Owners-generated content (Posts) in terms of metrics mentioned previously for the targeted community using Excel Graphics representation, in order to judge if these brands are well engaged or not we have proceed as follows :

On the one hand, we have chosen from each subcategory a well engaged brand in the world. The latter are selected using "Ranking the Brands" [4] statistics that shows the world most engaged brands at the same period when we have collected our data. In fact, we have made the same process of *Data Collection* and *Data Analysis* for these chosen brands to compare there results with those of our targeted community. Allows us to judge this fact, if the target community uses these metrics as conform as global standards.

On the other hand, we have calculated the average of engagement rate received by brands in term of each metric using the following formula :

$$EngagementRate = \frac{reactions + comments + shares}{posts} * \frac{100}{followers}$$

Where, $reactions$ is the total number of posts reaction including likes, loves, wows, hahas, sads and angrys. The $comments, shares, posts, followers$ are the number of comments, shares, posts, followers respectively.

Then, we confront the engagement rate with the Average number of posts published in term of the same metrics, to decide if the brand posts show optimal engagement.

4. Ranking the Brands : is a marketing agency that publishes yearly Top 100 lists, it is the central source for all published brand rankings. See "https ://www.rankingthebrands.com"

For some other metrics, we have relied on SocialBakers [5] studies and that of Buffer [6] and Buzzsumo [7] which is considered as one of the largest studies, where they analyzed 43 million Facebook posts from the top 20 000 brands in the world. [8].

#### 1.2.4 Customer-Generated Content Analysis

For this level, we analyse customers interest through different measures of interactions.

Concerning customer sentiments analysis, we have relied on Mataoui's et al. (2016) framework. In fact, we have improved and adapted it to fit with our case-study. They have proposed a lexicon-based process for vernacular Algerian Arabic sentiment analysis as mentioned in Chapter 2. We have used the following modules :

- **Preprocessing Module** : which is based on tokenization, normalization and stop-words removal steps that are implemented on a parser.
- **Language Detection and Stemming Module** : in which they used Khoja [9] stemmer as a light stemming Arabic tool to calculate the stems of tokens and some Google API for language detection and translation.
- **Polarity Computation Module** : in which they use the three following lexicons :
  - Keywords lexicon (L1) : which is composed of more than 3000 Arabic and ALGD words with their assigned polarities (positive/negative)
  - Negation words lexicon (L2) : it contain 54 negative words like ماشي، خاطي، لهلا ..الخ.
  - Intensification-words Lexicon (L3) : it contain 47 intensification words like جدا، للغايه، بزاف ..الخ.

.

In order to improve and adapt Mataoui's et al. work we have added an abbreviation treatments by prepared a list of about 100 abbreviations. In fact, we have extracted them from the comments dataset using some regular expressions then we have proceeded manually the resulting list and giving the suitable full form for each ones.
The abbreviated word list is in Appendices B.1 (see Appendice).
Concerning the treatment, we have maintained it to the Mataoui's et al. code in order to replace the abbreviations with its full form. In addition, we have enriched the lexicons by some frequent words that we notice their absence in the latter, and the emoticons list by those which belongs to our targeted social media platform.

#### 1.2.5 The Distribution of Languages and Scripts Used

In this aspect, we aim to know the most written language deployed by the targeted community. The study has twofold :

- Helping the brands in marketing getting closer to their customers through their understandable language.
- For future research in this area to focus their studies on the most used languages in that community.

In addition, we have analyzed the written scripts (Arabic, Latin or mixed) and the use of emoticons language.

---

5. www.socialbakers.com : social media analytics platform.
6. Buffer (https ://buffer.com/) : is a software company where they develop online applications intended for managing social media content.
7. BuzzSumo (https ://buzzsumo.com/) : one of the largest platforms that offers a powerful social media analytics tools.
8. See : https ://buffer.com/resources/facebook-marketing-strategy
9. See : http ://zeus.cs.pacificu.edu/shereen/research.htm

## 2 Experiments and Results

In order to evaluate the performance of our CBE measurement process, we have narrowed it on the case of Arabic customers and brands especially Algerian ones. In what follows, we describe and justify how the dataset is collected. Then, we will report results and comment them.

### 2.1 Data Collection

Concerning the *Data Collection* step mentioned On Section 1, we have insured that we get a most representative sample. Table 3.1 reports the final 50 Brands/Services that we have selected.

| Category | Subcategory | # | Illustration | Details |
|---|---|---|---|---|
| Brand | Appliance | 6 | Condor Electronics, LG Algerie, ENIEM, ENIE, Start-Light | LE |
| | | | Cobra Electronics | ME |
| | Beauty/Hygiene | 4 | Awane, Bimbies, Venus | LE |
| | | | finessecepro | - |
| | Beverage | 6 | Rouiba-Jus, Vita-Jus, Cevital-boissons, Ngaous | LE |
| | | | Aroma-Café | ME |
| | | | CAFE-Boukhari | SE |
| | Dairy | 3 | Soummam, FALAIT-Tartino, Berber-fromage | LE |
| | Electronics/Phone | 3 | Oppo Algerie | LE |
| | | | HuaweimobileDZ | ME |
| | | | SonymobileDZ | - |
| | Food | 6 | Benamor, Safina, Sim, CevitalCulinaire, Bimo | LE |
| | | | Jumbo | - |
| | Furniture | 2 | Sotrabois menuiserie d'art | ME |
| | | | Dz-meuble | - |
| | Household Goods | 4 | Aigle | LE |
| | | | Nassah, Force Xpress | ME |
| | | | El-Bahdjadetergents | SE |
| | Industrial | 4 | Imetal-SIDER EL-ADJAR, SNVI, ENAP | LE |
| | | | TEXALG ex. Sonitex | ME |
| | Industrial/Auto | 3 | Renault'DZ, Dacia'DZ, Peugeot'DZ | LE |
| Services | Accommodation | 3 | El-Djazair, ElAurassi | LE |
| | | | El Biar hotel | SE |
| | Telecommunication | 3 | Djezzy'DZ, Mobilis, Ooredoo'DZ | LE |
| | Transportation/Airlines | 2 | Air Algerie, Tassili Airlines | LE |
| | Web Service | 1 | Ouedkniss.com | - |

TABLE 3.1 – Details on Chosen Brands and Services (LE : Large Enterprise ME : Medium sized Enterprise SE : Small Enterprise)

Regarding the social media platform, our team has chosen **Facebook** in view of its popularity on the Algerian community. Figure 3.2 illustrate the social media platforms used by the Algerian community from January to October 2017, which is our data extraction period. These statistics are provided by Statcounter [10]. It shows that Facebook is the most used platform with 75.94% followed by Youtube and Twitter with 11.37% and 8.28% respectively.

In the literature, most studies on CBE collect the data in a period of 2-6 months. In our case, we extended the studied period to 10 months in order to get enough data because we have noticed that there are some pages that do not publish or interact for months.

We can ask why the data is old, it is due to Facebook company that have restricted their data access policy [11] on **April 2018** as an attempt to protect user informations. Thus, it was difficult for us to extract more data after those policy changes.

10. StatCounter is a web traffic analysis website (See : gs.statcounter.com).
11. See : https ://newsroom.fb.com/news/2018/04/restricting-data-access/

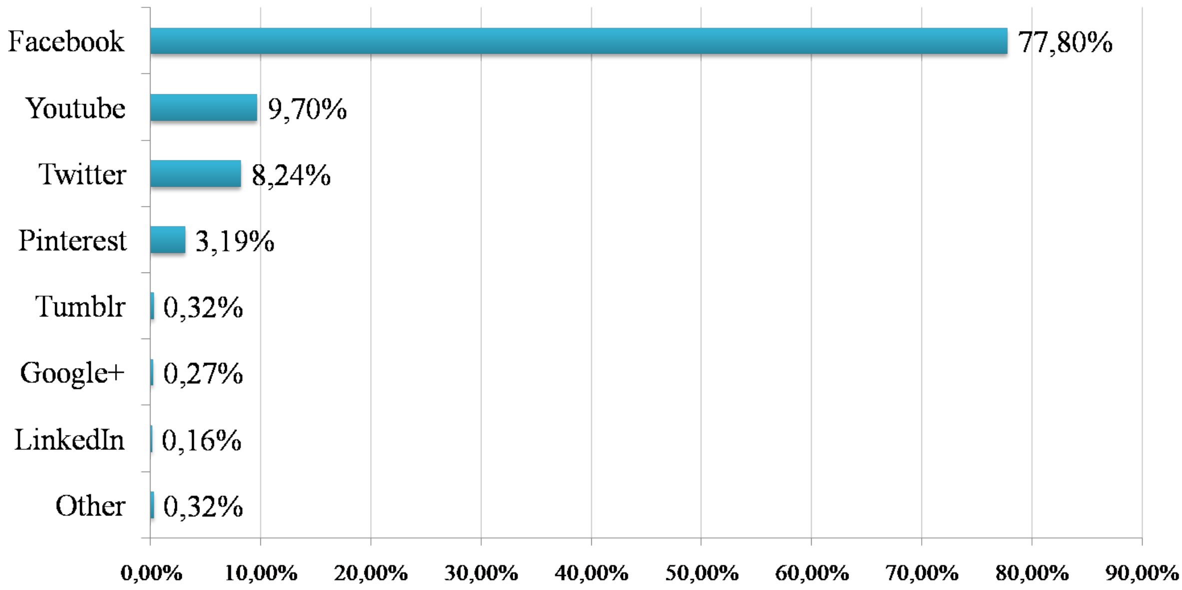



FIGURE 3.2 – Social Media Stats in Algeria in the period of January/2017 - October/2017 (StatCounter)

We have scraped all Facebook posts published by the selected brands/services between the period of January/2017 and October/2017. It is totalling more than 9 900 posts.

Let us mention that we have extracted posts with their related informations that may be essential for our study like number of reactions, comments and shares, we have also extracted all post's comments for 12 pages that belongs to different subcategories. In total, we obtain more than 800 000 comments. Table 3.2 shows general details about the extracted data and the performed annotations. For more details you can see Table A.1 and Table A.2 on the appendices A.

| Category | Subcategory | Illustration | Posts | | Comments | |
|---|---|---|---|---|---|---|
| | | | Scraped | Annotated | Scraped | Annotated |
| Brand | Appliance | Condor Electronics | 393 | 393 | - | - |
| | | ENIEM | 4 | 4 | - | - |
| | | ENIE | 29 | 29 | - | - |
| | | StartLight | 142 | 142 | - | - |
| | | LG Algerie | 261 | 261 | 21 786 | 5 999 |
| | | Cobra Electronics | 418 | 418 | - | - |
| | Beauty/ Hygiene | Awane | 227 | 227 | - | - |
| | | Bimbies | 183 | 183 | - | - |
| | | Venus | 127 | 127 | - | - |
| | | finessecepro | 49 | 40 | - | - |
| | Beverage | Rouiba-Jus | 114 | 114 | - | - |
| | | Vita-Jus | 0 | 0 | - | - |
| | | Cevital-boissons | 261 | 261 | - | - |
| | | Ngaous | 483 | 483 | - | - |
| | | Aroma-Café | 248 | 248 | 15 305 | 4 800 |
| | | CAFE-Boukhari | 0 | 0 | - | - |
| | Dairy | Soummam | 177 | 170 | - | - |
| | | FALAIT-Tartino | 95 | 95 | - | - |
| | | Berber-fromage | 158 | 158 | - | - |
| | Electronics/ Phone | Oppo Algerie | 453 | 453 | - | - |
| | | HuaweimobileDZ | 403 | 403 | - | - |
| | | SonymobileDZ | 81 | 81 | - | - |
| | Food | Benamor | 627 | 622 | - | - |
| | | Safina | 312 | 312 | - | - |
| | | Sim | 30 | 28 | - | - |
| | | CevitalCulinaire | 240 | 240 | - | - |

| Category | Subcategory | Illustration | Posts | | Comments | |
|---|---|---|---|---|---|---|
| | | | Scraped | Annotated | Scraped | Annotated |
| | | Bimo | 41 | 41 | - | - |
| | | Jumbo | 199 | 199 | - | - |
| | Furniture | Sotrabois menuiserie d'art | 129 | 129 | - | - |
| | | Dz-meuble | 78 | 70 | 31 398 | 6 599 |
| | Household Goods | Aigle | 185 | 185 | 11 281 | 7 063 |
| | | Nassah | 13 | 3 | - | - |
| | | Force Xpress | 235 | 235 | 26 576 | 6 000 |
| | | El-Bahdjadetergents | 12 | 12 | - | - |
| | Industrial | Imetal-SIDER EL-ADJAR | 39 | 39 | - | - |
| | | SNVI | 11 | 11 | - | - |
| | | ENAP | 30 | 30 | 256 | 162 |
| | | TEXALG ex. Sonitex | 1 | 1 | - | - |
| | Industrial/ Auto | Renault'DZ | 217 | 217 | - | - |
| | | Dacia'DZ | 213 | 213 | - | - |
| | | Peugeot'DZ | 278 | 278 | - | - |
| Services | Accommo-dation | El-Djazair | 27 | 25 | 42 | 32 |
| | | ElAurassi | 0 | 0 | - | - |
| | | El Biar hotel | 4 | 4 | 12 | 2 |
| | Telecomm-unication | Djezzy'DZ | 584 | 584 | - | - |
| | | Mobilis | 657 | 657 | 53 977 | 3 499 |
| | | Ooredoo'DZ | 719 | 719 | 565 284 | 5 000 |
| | Transporta-tion/Airlines | Air Algerie | 225 | 225 | 35 274 | 6 005 |
| | | Tassili Airlines | 32 | 32 | - | - |
| | Web Service | Ouedkniss.com | 565 | 565 | 45 539 | 4 840 |
| | | Total | 10 009 | 9 966 | 806 705 | 50 001 |

TABLE 3.2 – Details on the extracted and annotated dataset.

| Category | Subcategory | Illustration | Posts |
|---|---|---|---|
| Brand | Appliance | Beko | 99 |
| | Beauty/Hygiene | Colgate | 21 |
| | Beverage | CocaCola | 6 |
| | | Redbull | 65 |
| | Dairy | Danibio | 34 |
| | Electronics/Phone | Sumsung | 152 |
| | Food | Oreo | 2 |
| | | Reeses | 75 |
| | Industrial | Shell | 106 |
| | Industrial/Auto | BMW | 535 |
| Services | Web Service | Amazon | 575 |
| | | Total | 1 670 |

TABLE 3.3 – Details on the world extracted sample.

## 2.2 Data Analysis

Once the data set is collected, we have performed many preprocessings steps and annotations. According to their importance and feasibility, these annotations are done automatically, semi-automatically or completely manual.

### 2.2.1 Data Annotation

We have annotated more than 9 900 posts that belongs to 50 Algerian Brand/Service pages and 1 670 posts that belongs to 11 Brand/Service pages from the world sample. Figure 3.2 and Figure 3.3 shows general details about the Algerian and the world samples respectively.

Figure 3.4 shows an example of the performed manual annotation according to Content type.

TABLE 3.4 – Sample of posts annotation

| Posts | Content Type |
|---|---|
| شكراً لمحبين قهوة أروما ، إلتقط صورة مع أي من منتجات قهوة أروما وأرسلها لنا في صفحتنا لمشاركتها مع كل المعجبين في الشبكة الاجتماعية الفيسبوك. وأجمل صورة تفوز بهدية رمزية | Remuneration |
| ماهي الإختلافات بين الصّورتين؟ Quelles sont les différences entre les deux photos? | Entertainment |
| Le Stop Odeur ForceXoress a un effet duo; il cible les mauvaises odeurs pour les éliminer et empêcher leur réapparition. Il parfume l'air avec une fraîcheur qui duuuuure. | Information |

Concerning comments we have annotated 50 000 comments that belongs to over 80 posts taken from 12 Brand/Service pages, totaling more than 250 000 words. Table 3.5 illustrates how we have annotated them according to :

- Context : We have considered two sets, "**1**" which includes "In topic" comments that have relation with the targeted post or page and "**0**" which includes the "Out topic" comments that don't have any relation with the later.
- Polarity of Sentiment : We have considered three sets, "**-1**" which includes *negative* comments ,"**0**" for *neutral* comments and "**1**" for *positive* comments.
- Language Distribution : We have considered the most widely used languages for the Algerian community such as **MSA** which is the official language alongside the Tamazight one, the first foreign language **French**, the second foreign language **English**, and the **Algerian Dialect** as the common daily used language by the community. In fact, we have considered the ratio of words by language.

TABLE 3.5 – Sample of comment annotations

| Comment | In/Out | Polarity | MSA | Fr | Eng | Dialect |
|---|---|---|---|---|---|---|
| J'adore win yetba3ou | 1 | 1 | 0 | 1/3 | 0 | 2/3 |
| رايحين تهبلوني بمنتوجاتكم ، جربتهم كلهم روعة | 1 | 1 | 4/6 | 0 | 0 | 2/6 |

#### 2.2.2 Brand Owners-Generated Content Analysis

Let us recall that we have analyzed at this level the annotated posts in term of five metrics which are : Content Type (CT), Media Type (MT), Posting Time (PT), Frequency of Posting (FP) and Post Length (PL).

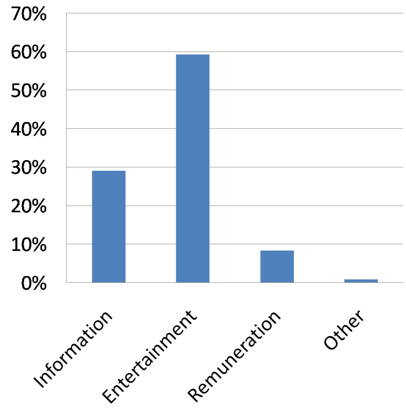
1. **Content Type** : Figure 3.3 reports the distribution of posts according to CT for both Algerian and the World samples and the related Engagement Rates.

   These results show that most of Algerian Brand/Service posts have Entertainment Content Type (ECT), and just 8% of them have Remuneration Content Type (RCT).

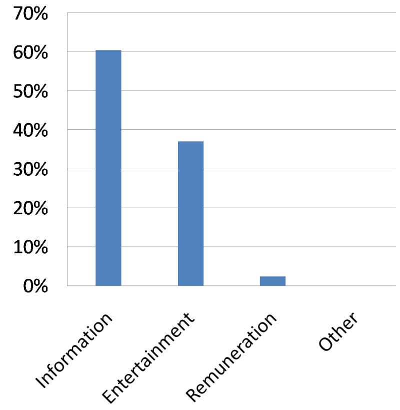
   In comparison of results shown in Figure 3.3a with those of Figure 3.3b, we notice that information posts are the most published posts in the World brands/services while they are less considerable in the Algerian ones.

   In addition, the comparison of results reported on Figure 3.3a with those Figure 3.3c, we notice that the most used CT is "*Entertainment*", while the "*Remuneration*" and the "*Information*" ones bring bigger ER than those of "*Entertainment*" in Algerian brand/service posts.

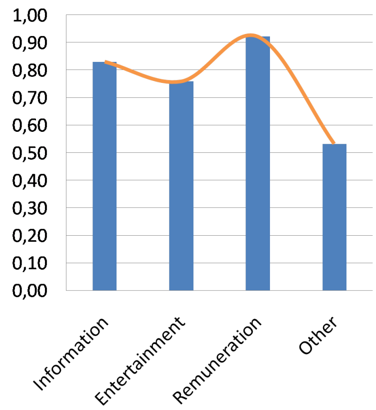
Comparing the curve of Algerian ER in Figure3.3c and the world's one in Figure3.3d, we notice that they have the same magnitude. In fact, the "*RCT*" is the most attractive CT followed by " *Information Content Type (ICT)*" for both Algerian and world post samples.



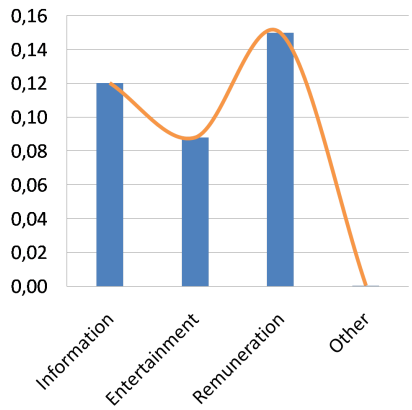

(a) Distribution of posts published by Algerian Brands/Services




(b) Distribution of posts published by World Brands/Services




(c) Average of ER received by Algerian Brands/Services




(d) Average of ER received by World Brands/Services

FIGURE 3.3 – Distribution of Posts & ER by CT (Algeria vs. World)

These results can help us to bring some recommendations to Algerian Brands/Services :

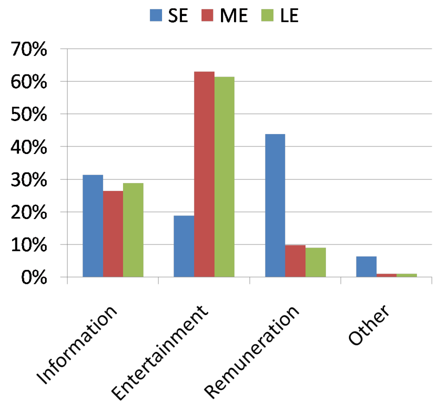
- Algerian post owners have to publish more informative content to get more engagement, and at the same time, this type of content gives them the opportunity to present more about their products/services to Facebook users.
- In addition, they must publish more Remuneration posts which are the most attractive content.

In addition, we have tried to analyze the distribution of Algerian Brand/Service posts in term of CT considering the company type. Figure 3.4 report these results. Surprisingly, the large companies investigate less in "*ICT*" and "*RCT*".

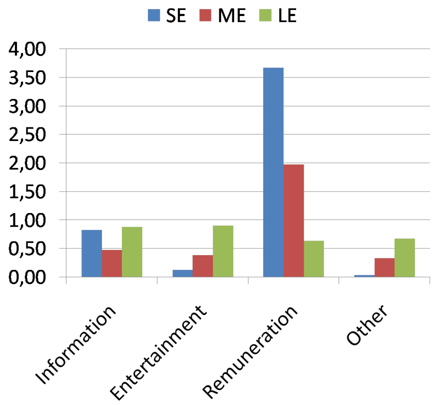
2. **Media Type** : Figure 3.5 reports the distribution of posts in term of MT for both Algerian and the world Brands/Services sample and their related Engagement Rates.

   These results show that the most used MT in Algerian Brand/Service posts are Photos.

   Comparing the results reported in Figure 3.5a with those of Figure 3.5b, we notice that Photo MT is the most used one for both Algerian and world samples, while video posts are less used on the Algerian Brand/Service posts comparing with the world one. This is due to cost of advertising video production.




(a) Distribution of posts by enterprise type




(b) Average of ER by enterprise type

FIGURE 3.4 – Distribution of Algerian Brand/Service Posts & ER by enterprise type in term of CT

According to comparative results of Figure 3.5a and Figure 3.5c, we notice that Photos, which are the most used MTs in Algerian brand/service posts, gives them the highest ER compared to others, concerning videos and status we notice that they bring a considerable ER while they are less used.

Contrasting the curve of Algerian ER in Figure 3.5c and the world ones in Figure3.5d, we observe that the link MT gives a considerable ER for the Algerian sample while this is not the case for the world sample. In counter part to the event MT which gives high ER for the world Brand/Service posts and low ER for the Algerian ones.

These results, can help us recommending to Algerian Brands/Services to Keep publishing photos and more video, status posts because they bring also a considerable ER.

In addition, we have analyzed the distribution of Algerian Brand/Service posts in term of MT considering the company type. Figure 3.6 reports these results. We notice that photo is the most used MT for all the company types, the link MT brings the biggest ER for ME.

3. **Posting Day And Time** : Figure 3.7 reports the distribution of Algerian Brand/Service posts in term of Posting Day (PD) and Posting Time (PT) and the related Engagement Rates.

   Concerning PD the results show that Algerian page owners publish more posts on *workdays* comparing with *weekend* days.

   According to comparative results (Figure 3.7a, Figure 3.7b), we notice that posts published on *workdays* bring more ER than those published on *weekend* days.

   Concerning PT, Figure 3.7c shows that Algerian page owners publish more posts around *11am* and *6pm*, while in contrast with Figure 3.7d we notice that posts published around *1pm*, *7pm*, *8pm* and *11pm* give more ER than the others.

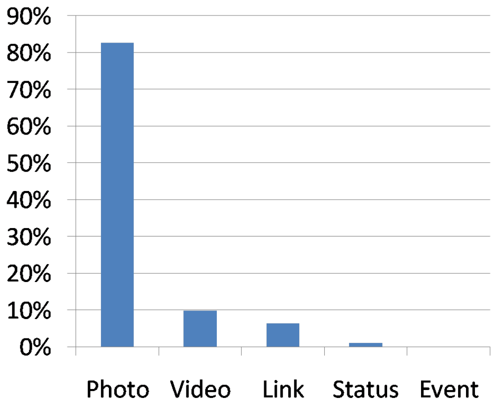
   Based on these findings, our suggested recommendations for the Algerian page owners are :

   - Keep on publishing more posts on workdays.
   - Publish more posts around *1pm*, *7pm*, *8pm* and *11pm* in order to get more ER.

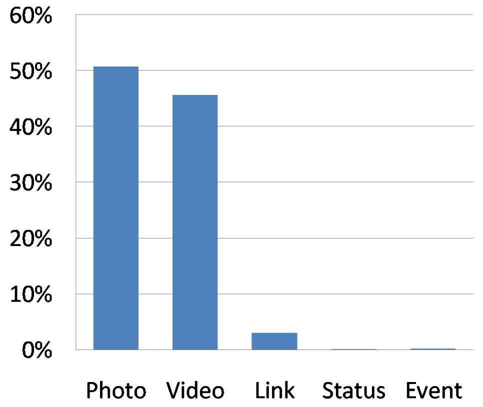
4. **Frequency of Posting** :

   By analyzing our sample dataset over a period of 10 months, we have found that :

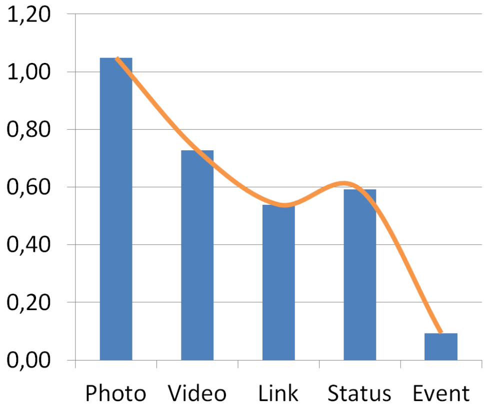
   - Algerian Brand/Service page owners post in a frequency of **0.7 Post/Day** which is about **5 Posts/Week**.



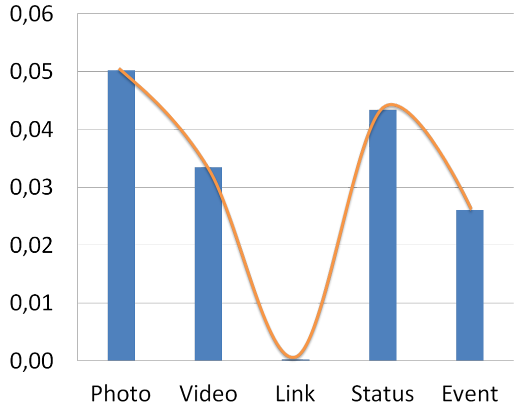

(a) Distribution of posts published by Algerian Brands/Services




(b) Distribution of posts published by World Brands/Services

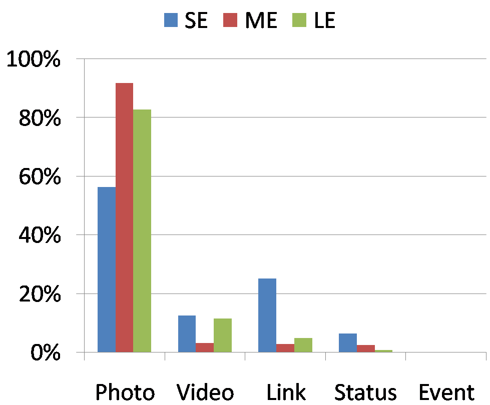



(c) Average of ER received by Algerian Brands/Services

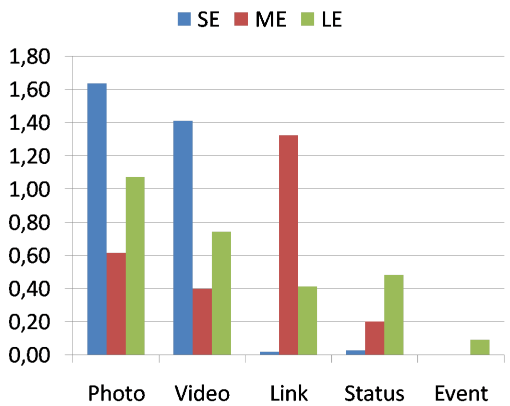



(d) Average of ER received by World Brands/Services

FIGURE 3.5 – Distribution of Posts & ER by MT (Algeria vs. World)




(a) Distribution of posts

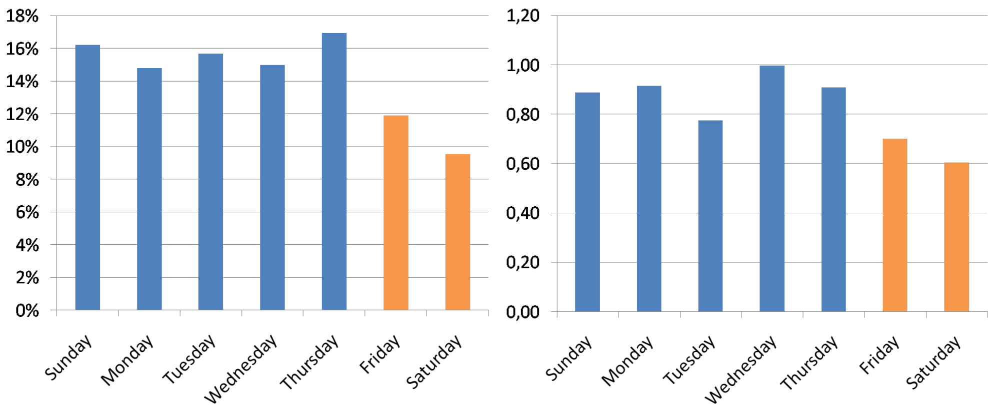



(b) Average of ER

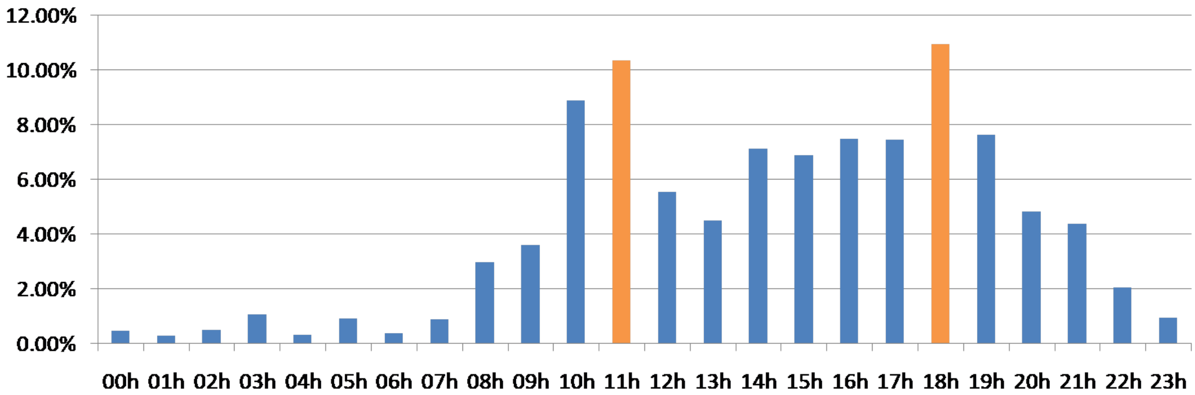
FIGURE 3.6 – Distribution of Algerian Brand/Service Posts & ER by enterprise type in term of MT

- The world Brand/Service page owners post in a frequency of **0.5 Post/Day** which is **4 Posts/Week** approximately.

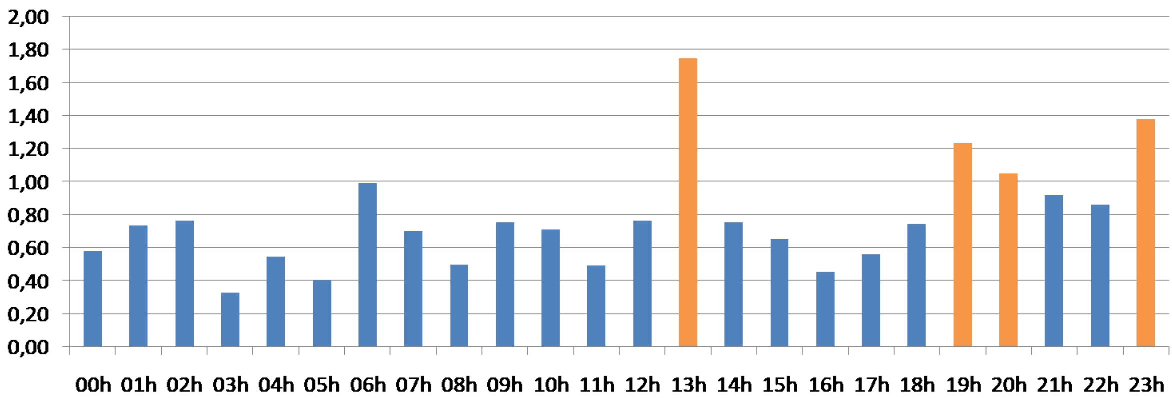
Basing on these results, we notice that the frequency of posting for the Algerian sample is better than that of the world one. While basing on **Socialbakers** study where they found that, according to their data, "higher publishing frequency on Facebook = higher




(a) Distribution of posts by PD (b) Average of ER by PD




(c) Distribution of posts by PT




(d) Average of ER by PT

FIGURE 3.7 – Distribution of Algerian Brand/Service Posts & ER in term of PD & PT

engagement". In addition according to **Buffer** and **buzzsumo** study we have found that "five posts a day seem optimal for engagement. But its possible to post more or less", we notice that the Algerian frequency is less considerable.

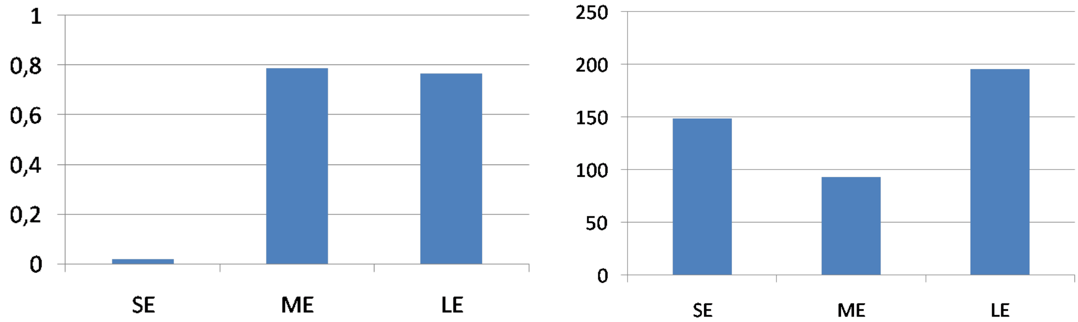
Figure 3.8a illustrates the distribution of the Frequency of Posting (FP) according to company type, we notice that SE have a very small frequency.

These results help us to recommend to Algerian Brand/Service page owners to increase their frequency of posting especially for the SE.

5. **Post Length** : By analyzing our sample dataset, we have found that Algerian Brand/Service posts have an average of **186 character** by post. While, World

Brand/Service posts have an average of **158 character** by post. These results have the same magnitude.

Figure 3.8b illustrate the average of post length according to the company types, we notice that the shortest posts are published by the ME.




(a) Frequency of posting

(b) Average of post length

FIGURE 3.8 – Average of post length & Frequency of posting in term of company type

Basing on a **BuzzSumo** study, Figure 3.9 shows the distribution of post length in term of average engagement, it reports that shortest posts bring more engagement. In contrast, we notice that Algerian Brand/service posts are so longer.

Thus, our recommendation to Algerian brand/service page owners is to reduce the number of characters within a post as mush as possible.

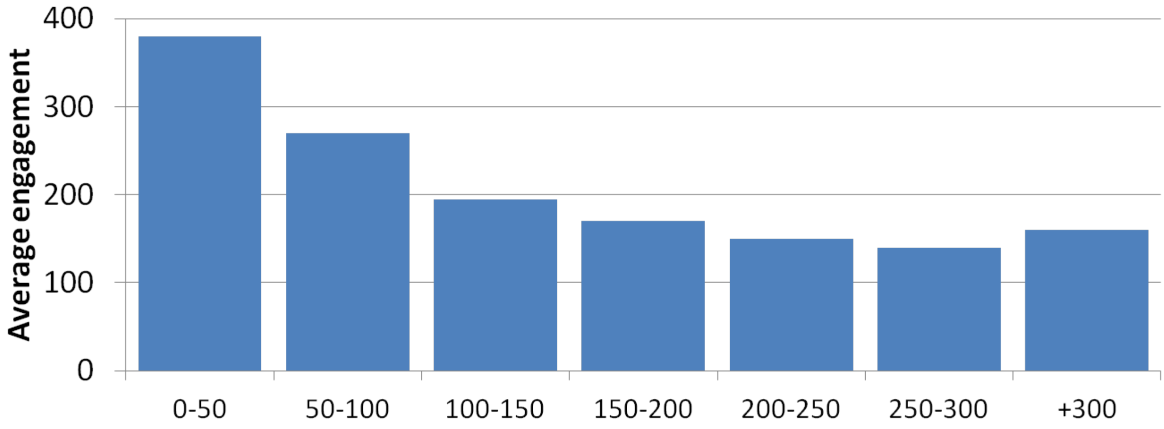



FIGURE 3.9 – Number of Characters vs. Average Engagement ((Moeller, 2019))

### 2.2.3 Customers-Generated Content Analysis

Let us recall that we have opted to analyze Algerian CGC (likes, comments and shares) in Facebook. Figure 3.10 give us a general view on the *interactions* and the *emotional reactions* for the Algerian customers.

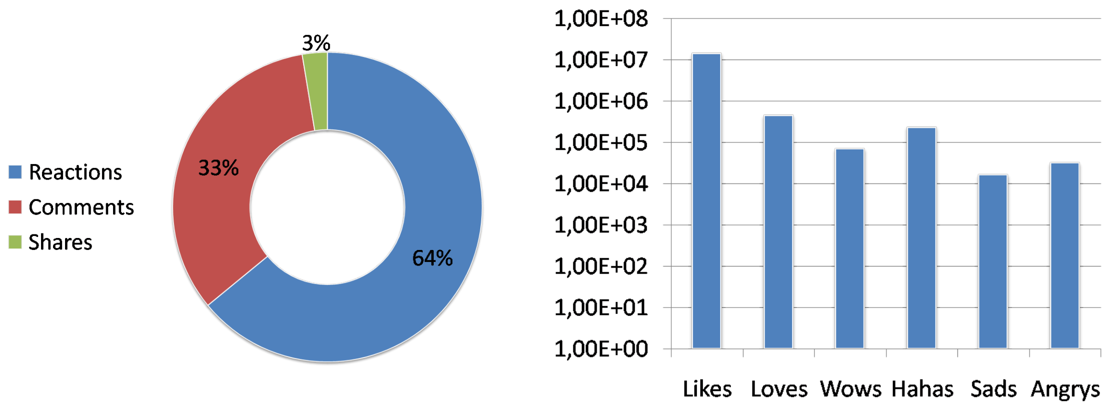
The "Reactions" set includes emotional reactions : likes, loves, wows, hahas, Sads and angrys.

Figure 3.10b shows that for the Algerian customers the *emotional reactions* are the most used interaction type followed by *comments*, while the *shares* are less considered. In addition, Figure 3.10a reports that customers mainly use Like reactions, while they are less used to the other emotional types. It can be understand that , in 2017 these social signals functionality are recently added by Facebook.

Figure 3.11 illustrates the distribution of CGC according to **Like Ratio (LR)**, **Comment Ratio (CR)**, and **Share Ratio (SR)** mentioned on Section 1.2.1 for both Algerian and world customers.

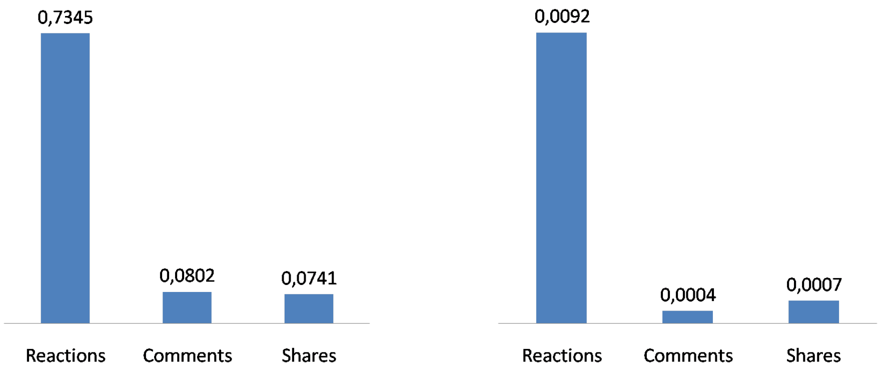
According to comparative results (Figure 3.11a, Figure 3.11b), we observe that both Algerian and world sample customers mainly use *emotional reactions*, while *shares* are less used by Algerian customers than *comments* unlike the world customers who use *shares* more than *comments*.

Figure 3.12 reports the average number of interaction per week according to the Algerian and the World samples. We observe that the Algerian customers are less active on reacting and sharing than the world ones, while concerning the comments they are more active.




(a) Distribution by type of Social Signal (b) Distribution by type of emotional Reactions

FIGURE 3.10 – Distribution of CGC by Social signals and Emotional Reactions (Algeria)




(a) Distribution Algerian Customer Interaction Ratios (b) Distribution World Customer Interaction Ratios

FIGURE 3.11 – Distribution of CGC by by interaction ratios (Algeria vs. World)

1. **Interaction Duration** Concerning this metric we have found that Algerian brand/service posts take in average 40 days as ID. Figure 3.13 illustrates the distribution of ID according to the selected sample. It shows that post ID very variable according to the brand.
2. **Sentiment Analysis** In order to measure Algerian customer sentiments, we have analyzed the annotated comments. Figure 3.14 reports the distribution of posts according to

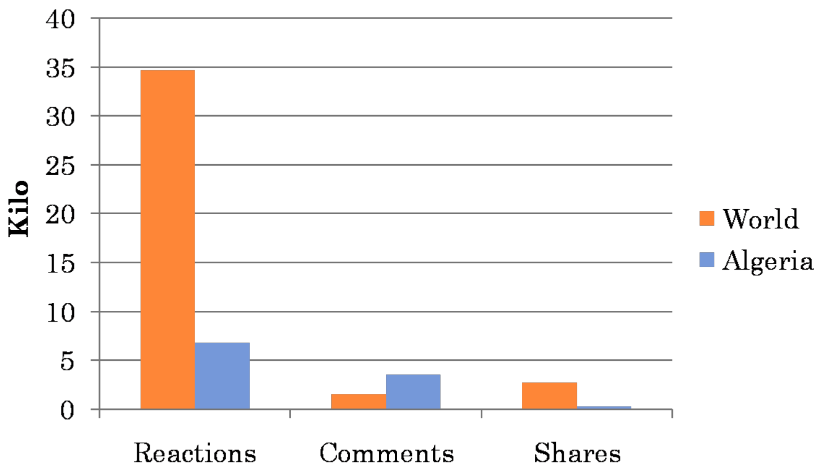



FIGURE 3.12 – Average number of interaction per week according to the Algerian and the World samples

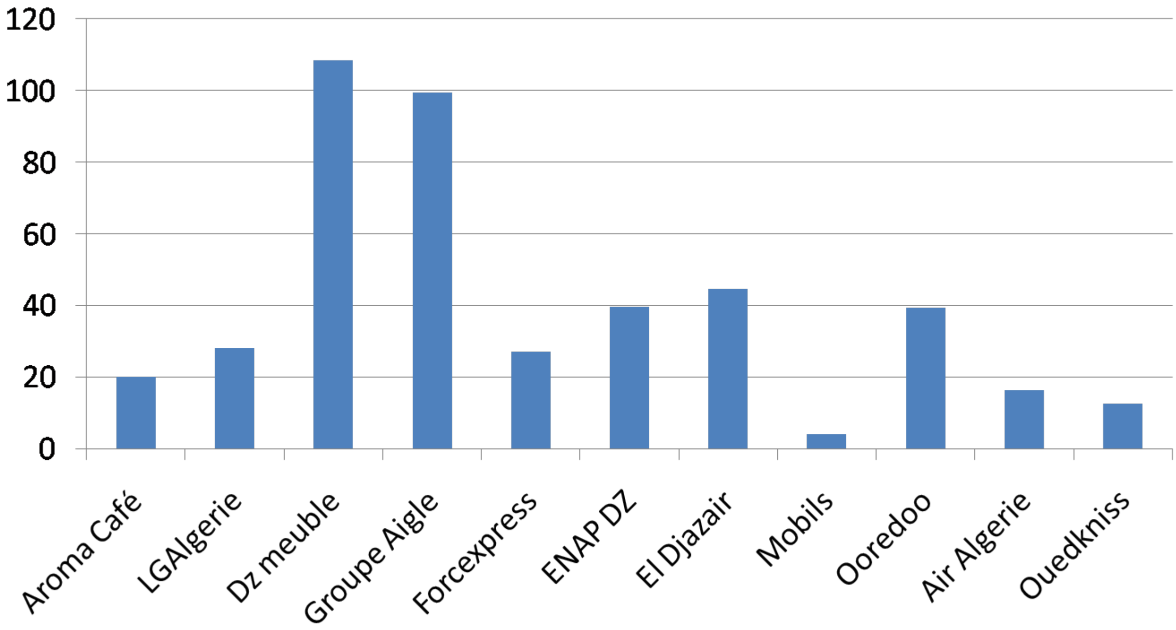


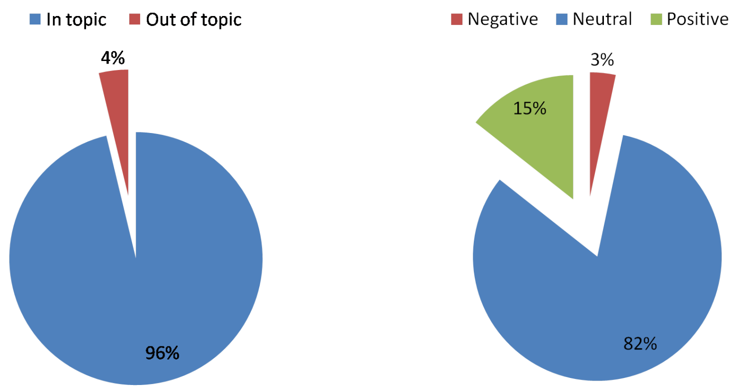

FIGURE 3.13 – Distribution of interaction duration by Brand/Service pages

their context if it has any relation with the targeted page or post, and the distribution of these comments in term of their polarity.

Figure 3.14a shows that most of the annotated comments tackled the same topic with their related post or page and just 4% of them are totally out of topic. These comments (the "In topic" set) are partitioned on three other sets according to their polarity as its shown in Figure 3.14b. Thus, we notice that most comments are neutral and that the percentage of positive comments is greater than the negative ones.

Considering just the positive and negative comments, Figure 3.15 reports some details on the distribution of comments polarity overs the analyzed brands/services.

We observe that comments belongs to **Ouedkniss**, **Mobilis**, **Ooridoo**, and **Airalgerie** pages have a big amount of negativity. Thus, we recommend to those brands/services to change their social media strategies in order to achieve their customers satisfaction.




(a) Distribution of comments context (b) Distribution of comments polarity

FIGURE 3.14 – Distribution of Algerian customers comment by context and polarity

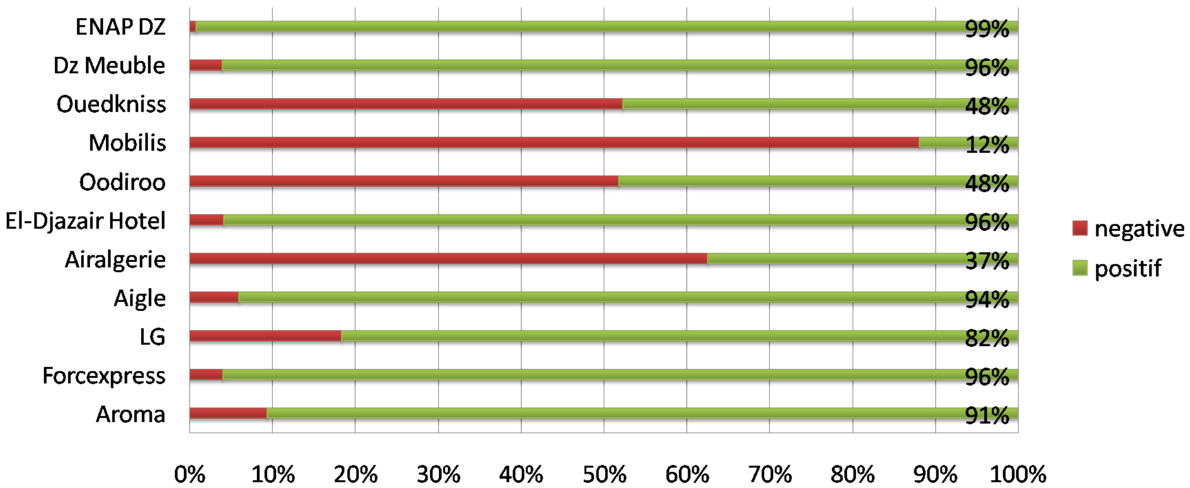



FIGURE 3.15 – Distribution of comments polarity by Brand/Service pages

### 2.2.4 The Distribution of Used Languages and Scripts Analysis

Concerning the used languages, Figure 3.16 illustrates their distribution in customer comments. It shows that French is the most used language on Algerian customer's comments followed by Arabic with 35.7% then Algerian dialect with 19.4%. The category "Other" includes languages like Tamazight, Espagnol, Korean and others.

Figures 3.17 and 3.18 give more details on the distribution of languages by brands/services and by subcategories.

In addition, we have studied the distribution of scripts used by Algerian customers including *Arabic scripts*, *Latin scripts*, *mixed* Arabic and Latin scripts and an *other* scripting set like emoticons or numbers. Figure 3.19a shows these statistics. We observe that 65% of textual comments use Latin characters while 25% of them use Arabic ones.

Further, we have analyzed the usage of *Emoticons Language*, Figure 3.19b shows that just 7% of Algerian customer comments are using emoticons while 93% of them are textual.

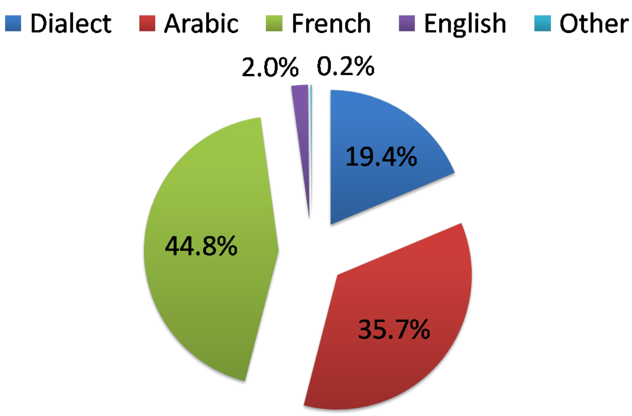



FIGURE 3.16 – Distribution of used languages in comments

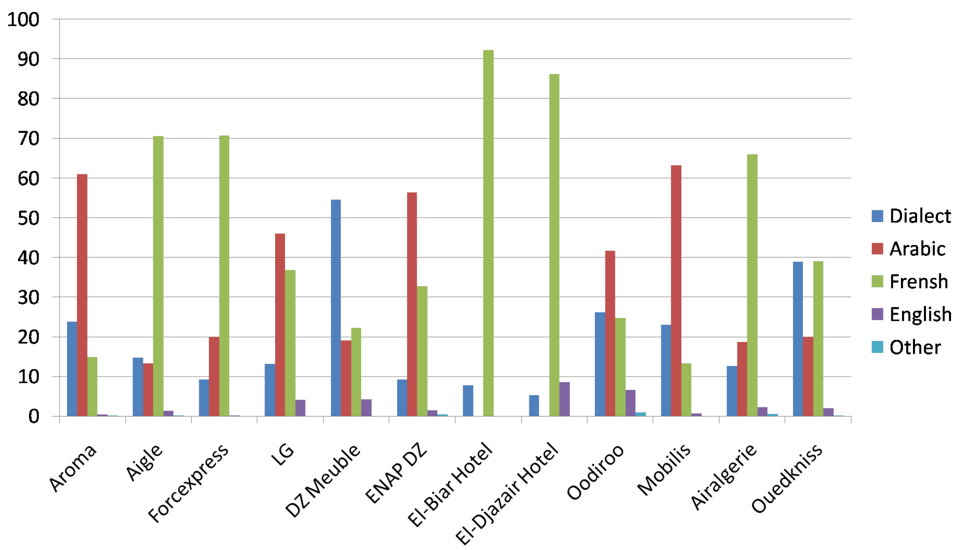



FIGURE 3.17 – Distribution of languages by pages

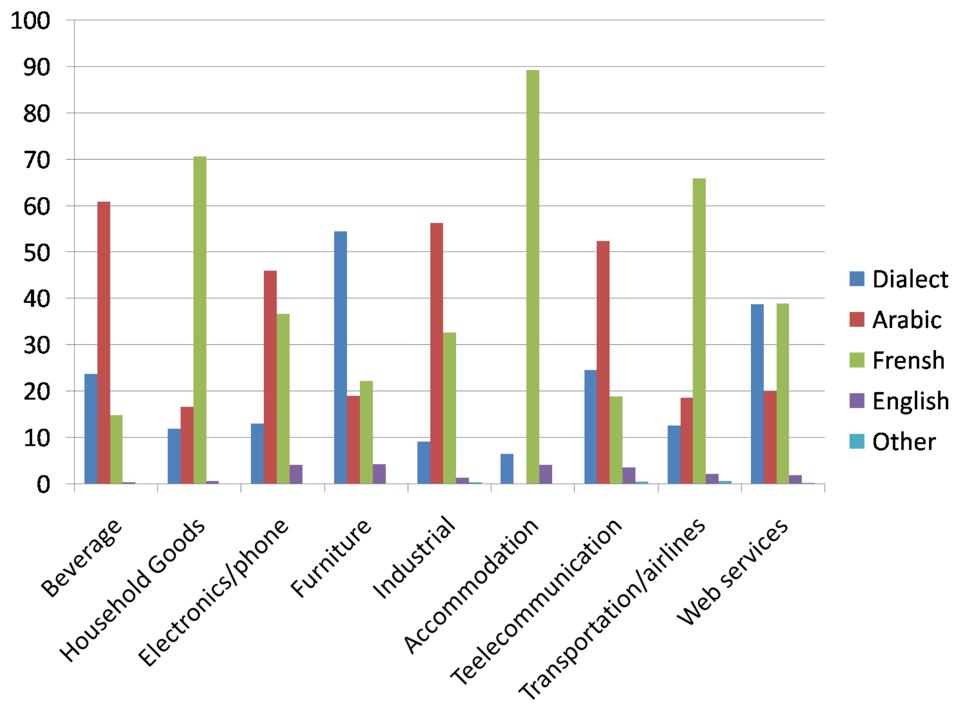
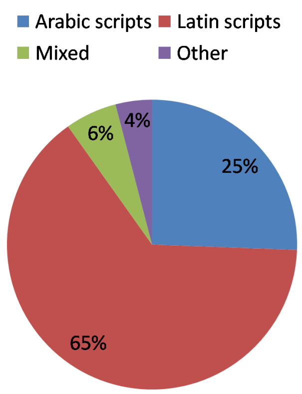
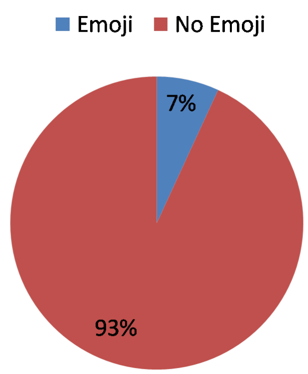



FIGURE 3.18 – Distribution of languages by subcategory




(a) Scripts




(b) Emojies

FIGURE 3.19 – Distribution of used scripts and emoticons on Algerian customer comments

## 3 Main Findings and Recommendations

From our experiments, we have derived some conclusions concerning the usage of social media signals by Algerian brands/services and their users/customers.

Among the main findings;

1. The brand owners post less attractive content according to their customers. In fact, they are mainly publishing entertainment content, while the informative and the remuneration ones are less considered.
2. Concerning the media type, they use mainly photos which provide a considered engagement in comparison with the other medias.
3. They post more on workdays and around 11h and 18h.
4. In general, they are quiet engaged as they publish in average 5 posts per week which is considered as lower bound to consider the brand as engaged. However the SE category has a very low frequency of posting.
5. Most comments are neutral.
6. Some pages were able to reach the satisfaction of their customers, while other are not yet.
7. Algerian customers use mainly *French* words in their comments, this confirm the existence of code switching phenomenon in Algerian daily communications.

On the light of these results, we can make some useful recommendations to Algerian brands/services when they deal with SMM :

- Algerian post owners should pay more attention to the informative and the remuneration content.
- They should continue to post photos, but without neglecting other medias like video, status and links.
- They should focus on publishing on workdays, at 11h and the period from 19h to 23h.
- They have to publish at least one post per day with a maximum of 50 characters to get more users engagement.
- Pages where customers are not satisfied, must change their marketing strategies.

## 4 Conclusion

In this chapter, we have devised an approach that can bring a detailed overview in social media and the behavior of their users/customers. This approach is applied on Algerian brands/services on Facebook platform. These measurements take into account many influencing factors and metrics. Main findings are outlined and some recommendations are made to that effect.

# Conclusion and Future Work

In our research, we have aimed to devise an approach that gives some information on the engagement of Algerian brands and services as on the user/customer engagement.

The extraction of such information from the brand side gives or decides whether Algerian brands are aware of the SMM usage. From the other side, it measures the customer engagement to extract information about their behaviors on the SM and the level of their satisfaction about brand products.

The proposed approach is based on two mechanisms. First, it deployed engagement metrics such engagement rate metric, CT, MT, etc. The second mechanism is the analysis of the user generated content. This approach is narrowed on the Facebook brand pages, their posts and their related comments and reactions. Our measurement process concluded regarding Algerian brands that they are well engaged compared to the world ones in terms of some used metrics. But this does not prevent us from saying that in terms of other metrics, Algerian brands are not at the required level in contrast with world brands/services. Regarding the CE and concerning the sentiment analysis part, Algerian user content are in general more objective, in some brands they are totally negative, but the positive comment rate is bigger than the negative one.
These results help us to propose some recommendations that upgrade the engagement level more and enabling companies to adopt new strategies to avoid shortcomings in the previous ones.

This research aimed to measure the engagement in social media in case of the Algerian brands and services, this analysis include two levels : the brand engagement from a side for the reason to conclude whether Algerian brands are aware of the SMM usage and if they're well engaged or not, and from the other side to measure the customer engagement to extract information about their behaviors on the SM and the level of their satisfaction around a targeted product, the results presented in this paper are limited to Facebook brand pages (comment, post) as SMM platform.

Our measurement process concluded regarding Algerian brands that they are well engaged compared to the world one in terms of some used metrics, but this does not prevent us from saying that in terms of other metrics, Algerian brands are not at the required level in contrast with world brands/services. Regarding the Customer engagement and concerning the sentiment analysis part, Algerian customers are in general neutral, in some brands they are totally negative, but the positive comment rate is bigger than the negative one.

These results help us to propose some recommendations that upgrade the engagement level more and enabling companies to adopt new strategies to avoid shortcomings in the previous ones.

In order to confirm our findings or to understand other different behaviors in SM, we plan to enlarge our proposed process to deal with Hashtags analysis to measure the engagement, more content analysis for both posts and comments. In fact, we will define some new lexicons that improve the engagement in addition, let mention that our analysis has focused on Facebook pages. So in the future work, we plan to deal with other social media platforms.

# References


Barger, V., Peltier, J. W., and Schultz, D. E. (2016). Social media and consumer engagement : a review and research agenda. *Journal of Research in Interactive Marketing*, 10(4) :268–287.

Bartoletti, M. (2013). Bachelor of Business Administration in Tourism and Hospitality Management. page 66.

Batrinca, B. and Treleaven, P. C. (2015). Social media analytics : a survey of techniques, tools and platforms. *AI & SOCIETY*, 30(1) :89–116.

boyd, d. m. and Ellison, N. B. (2007). Social Network Sites : Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1) :210–230.

Dessart, L. (2017). Social media engagement : a model of antecedents and relational outcomes. *Journal of Marketing Management*, pages 1–25.

Drury, G. (2008). Opinion piece : Social media : Should marketers engage and how can it be done effectively ? *Journal of Direct, Data and Digital Marketing Practice*, 9(3) :274–277.

Fan, W. and Gordon, M. D. (2014). The power of social media analytics. *Communications of the ACM*, 57(6) :74–81.

Gonçalves, A. (2017). *Social Media Analytics Strategy : Using Data to Optimize Business Performance*.

Guellil, I., Adeel, A., Azouaou, F., and Hussain, A. (2018). SentiALG : Automated Corpus Annotation for Algerian Sentiment Analysis. *arXiv :1808.05079 [cs]*. arXiv : 1808.05079.

Harvey, K. (2014a). Social Media, Definition and Classes of. In *Encyclopedia of Social Media and Politics*, pages 1158–1161. SAGE Publications, Inc., Thousand Oaks.

Harvey, K. (2014b). User-Generated Content. In *Encyclopedia of Social Media and Politics*, pages 1313–1314. SAGE Publications, Inc., Thousand Oaks.

Hollebeek, L. D. (2011). Demystifying customer brand engagement : Exploring the loyalty nexus. *Journal of Marketing Management*, 27(7-8) :785–807.

Jayasingh, S. and Venkatesh, R. (2015). Customer Engagement Factors in Facebook Brand Pages. *Asian Social Science*, 11(26).

Jurafsky, D. and Martin, J. H. (2009). *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall series in artificial intelligence. Prentice Hall, Pearson Education Internat, Upper Saddle River, NJ, 2. ed., pearson internat. ed edition.

Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite ! The challenges and opportunities of Social Media. *Business Horizons*, 53(1) :59–68.

Karjaluoto, H., Munnukka, J., and Tiensuu, S. (2015). The Effects of Brand Engagement in Social Media on Share of Wallet. page 15.

Kumar, E. and I. K. International Publishing House (2012). *Natural language processing*. I. K. International Publishing House Pvt. Ltd., New Delhi ; Bangalore.

Lalmas, M., O'Brien, H., and Yom-Tov, E. (2014). Measuring user engagement. 6(4) :1–132.

Leckie, C., Nyadzayo, M. W., and Johnson, L. W. (2016). Antecedents of consumer brand engagement and brand loyalty. *Journal of Marketing Management*, 32(5-6) :558–578.

Martin, J. (2004). Study and Supervision Guide. page 80.

Mataoui, M., Zelmati, O., and Boumechache, M. (2016). A Proposed Lexicon-Based Sentiment Analysis Approach for the Vernacular Algerian Arabic. *Research in Computing Science*, 110 :55–70.

McAlexander, J. H., Schouten, J. W., and Koenig, H. F. (2002). Building Brand Community. *Journal of Marketing*, 66(1) :38–54.

Moeller, S. (2019). The 2019 Ultimate Guide to Facebook Engagement.

Muniz, A. M. and O'Guinn, T. C. (2001). Brand Community. *Journal of Consumer Research*, 27(4) :412–432.

Nadaraja, R. and Yazdanifard, R. (2013). Social Media Marketing : Advantages and disadvanteges. *Social Media Marketing*, page 11.

Olczak, A. B. and Sobczyk, R. K. (2013). Brand Engagement on Facebook Based on Mobile Phone Operators' Activity Within Four European Countries. *Studia Ekonomiczne*, 150 :97–108.

Perreault, M.-C. and Mosconi, E. (2018). Social Media Engagement : Content Strategy and Metrics Research Opportunities.

Phillips, J. (2014). *Digital analytics primer*. Pearson, Upper Saddle River, N.J.

Pletikosa Cvijikj, I. and Michahelles, F. (2013). Online engagement factors on Facebook brand pages. *Social Network Analysis and Mining*, 3(4) :843–861.

Quesenberry, K. A. and Coolsen, M. K. (2018). What Makes Facebook Brand Posts Engaging ? A Content Analysis of Facebook Brand Post Text That Increases Shares, Likes, and Comments to Influence Organic Viral Reach. *Journal of Current Issues & Research in Advertising*, pages 1–16.

Rahab, H., Zitouni, A., and Djoudi, M. (2017). ARAACOM : ARAbic Algerian Corpus for Opinion Mining. In *Proceedings of the International Conference on Computing for Engineering and Sciences - ICCES '17*, pages 35–39, Istanbul, Turkey. ACM Press.

Saragih, M. H. and Girsang, A. S. (2017). Sentiment analysis of customer engagement on social media in transport online. In *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, pages 24–29, Malang. IEEE.

Sherman, A. and Smith, D. E. (2013). *Social media engagement for dummies*. For dummies. John Wiley & Sons Inc, Hoboken, N.J.

Soumeur, A., Mokdadi, M., Guessoum, A., and Daoud, A. (2018). Sentiment Analysis of Users on Social Networks : Overcoming the challenge of the Loose Usages of the Algerian Dialect. *Procedia Computer Science*, 142 :26–37.

Tsagkias, M. (2012). *Mining social media : tracking content and predicting behavior.* PhD thesis, SIKS, Amsterdam.

# Appendices

# Annexe A

# Details on the Scraped Dataset

## Details on the Algerian Scraped Dataset

| Category | Subcategory | Illustration | Funs | Posts | Reactios | Comments | Shares |
|---|---|---|---|---|---|---|---|
| Brand | Appliance | Condor Electronics | 788651 | 393 | 488222 | 35917 | 14337 |
| | | COBRA Electronics | 13884 | 418 | 14103 | 3615 | 3635 |
| | | ENIE | 4713 | 29 | 1253 | 257 | 253 |
| | | ENIEM | 5589 | 4 | 651 | 114 | 78 |
| | | LGAlgerie | 569825 | 261 | 152158 | 15832 | 12904 |
| | | Starlight | 20906 | 142 | 24178 | 1764 | 826 |
| | Beauty/ Hygienne | Awane Faderco | 511173 | 227 | 193670 | 9818 | 2359 |
| | | BimbiesFaderco | 205881 | 183 | 142099 | 4693 | 2502 |
| | | Finessecepro | 102487 | 40 | 120378 | 1763 | 550 |
| | | Laboratoires VENUS SAPECO | 356461 | 127 | 31291 | 4407 | 664 |
| | Beverage | Aroma Café Officiel | 157889 | 248 | 180926 | 13368 | 5617 |
| | | Café boukhari | 1260 | 0 | 0 | 0 | 0 |
| | | Cevital Agro-Industrie boissons | 641567 | 261 | 443233 | 196300 | 11582 |
| | | Rouiba Jus | 1516712 | 114 | 621681 | 16062 | 5163 |
| | | Ngaous | 114380 | 483 | 88593 | 4818 | 1345 |
| | | VITAJUS | 69177 | 0 | 0 | 0 | 0 |
| | Dairy | Laiterie Soummam | 28018 | 170 | 28606 | 2863 | 2607 |
| | | Berber fromage SARL PROMASIDOR DJA-ZAIR | 345882 | 158 | 49447 | 6652 | 5249 |
| | | Sarl FALAIT Tartino Fromage | 64411 | 95 | 5720 | 417 | 281 |
| | Electronics /Phone | OPPOAlgerie | 33675405 | 453 | 813793 | 28939 | 22744 |
| | | Huawei mobile DZ | - | 403 | 769930 | 27352 | 11913 |
| | | Sony mobile dz | - | 81 | 463411 | 10584 | 6489 |
| | Food | Groupe Bimo Algerie | 97668 | 41 | 37873 | 2309 | 697 |
| | | Amor Benamor | 2744546 | 622 | 954494 | 86972 | 62262 |
| | | Jumbo Algérie | 188077 | 199 | 203357 | 16418 | 4934 |
| | | Safinadz | 800513 | 312 | 570755 | 48792 | 116654 |
| | | Groupe Sim | 4423 | 28 | 1473 | 123 | 67 |
| | | Cevital Agro-Industrie | 870130 | 240 | 1089043 | 46870 | 31590 |
| | Furniture | Dz meuble | 714155 | 70 | 318561 | 28309 | 12422 |
| | | Sotrabois menuiserie d'art | 112629 | 129 | 92894 | 4515 | 3629 |
| | Household Goods | Groupe Aigle | 92222 | 185 | 49239 | 14091 | 5156 |
| | | GROUPE NASSAH | 1039 | 3 | 32 | 4 | 0 |
| | | Forcexpress | 100899 | 235 | 59460 | 16693 | 2564 |
| | | El-Bahdja Détergents | 96480 | 12 | 27176 | 188 | 150 |
| | Industrial | entreprise Nationale des Peintures (ENAP) | 916 | 30 | 1384 | 183 | 331 |

| Category | Subcategory | Illustration | Funs | Posts | Reactios | Comments | Shares |
|---|---|---|---|---|---|---|---|
| | | Groupe SNVI ex. Sonacom | 4185 | 11 | 1111 | 81 | 72 |
| | | TEXALG ex. Sonitex Cotitex | 448 | 1 | 0 | 0 | 0 |
| | | Imetal SIDER EL HADJAR EPE | 1497 | 39 | 2130 | 321 | 447 |
| | Industrial/ Auto | Renault Algerie | 1197791 | 217 | 1186911 | 60277 | 24125 |
| | | Dacia Algerie | 539651 | 213 | 583482 | 38446 | 6231 |
| | | peugeot algerie | - | 278 | 585505 | 14175 | 11094 |
| Services | Accommendation | ElBiar.Hotel | 20959 | 4 | 33 | 0 | 0 |
| | | El-Aurassi Hotel | 9940 | 0 | 0 | 0 | 0 |
| | | Chaine El Djazair | 1387 | 25 | 303 | 29 | 46 |
| | Telecommunication | Djezzy | 4008451 | 584 | 1078914 | 134215 | 43641 |
| | | Mobils | 2653947 | 657 | 945153 | 159693 | 20016 |
| | | Ooredoo | 5092218 | 719 | 1844904 | 6534648 | 90568 |
| | Transportation/Airlines | Tassili Airlines | 23074 | 32 | 2595 | 432 | 247 |
| | | Air Algerie | 362817 | 225 | 106416 | 17453 | 5254 |
| | Web Sevices | Ouedkniss | 1412396 | 565 | 314631 | 33529 | 46752 |
| | | Average | 1 283 973 | 199 | 293 823 | 152 886 | 12 001 |

TABLE A.1 – Details on the Algerian dataset.

# Details on the World Scraped Dataset

| Category | Subcategory | Illustration | Funs | Posts | Reactios | Comments | Shares |
|---|---|---|---|---|---|---|---|
| Brand | Appliance | Beko | 5657932 | 99 | 361185 | 13065 | 8208 |
| | Beauty/Hygiene | Colgate | 3665366 | 21 | 230953 | 5997 | 6659 |
| | Beverage | CocaCola | 107155366 | 6 | 2340 | 378 | 99 |
| | | Redbull | 48908658 | 65 | 402905 | 14875 | 53295 |
| | Dairy | Danibio | 1881520 | 34 | 40631 | 914 | 1107 |
| | Electronics/Phone | Sumsung | 43401587 | 152 | 4310969 | 185352 | 279797 |
| | Food | Oreo | 43115607 | 2 | 3880 | 132 | 177 |
| | | Reeses | 12166480 | 75 | 1774074 | 234513 | 403905 |
| | Industrial | Shell | 7593853 | 106 | 1650057 | 37983 | 71719 |
| | Industrial/Auto | BMW | 20213671 | 535 | 7171881 | 129403 | 320907 |
| Services | Web Service | Amazon | 28464072 | 575 | 631545 | 107871 | 157255 |
| | | Average | 29 293 101 | 152 | 1 507 311 | 66 408 | 118 466 |

TABLE A.2 – Details on the World dataset.

# Annexe B

# List of Common Abreviation Word

| Short Form | Full Form | Short Form | Full Form | Short Form | Full Form |
|---|---|---|---|---|---|
| jdr | j'adore | 2m1 | demain | 2r1 | de rien |
| att | attend | avc | avec | b1 | bien |
| bcp | beaucoup | bn8 | bonne nuit | Bnj | bonjour |
| Bns | bonsoir | brk | bark | bro | frère |
| bsr | bonsoir | bzf | bezaf | c | c'est |
| cad | c'est-à-dire | cbn | c'est bon | cc | coucou |
| cest | c'est | cpg | C'est pas grave | ctd | c'est-à-dire |
| ctt | c'est tout | cv | ça va | d1 | demain |
| da | dinar | d'ac | d'accord | dak | d'accord |
| dcr | d'accord | dqp | Dès que possible | dr1 | de rien |
| dsl | désolé | dz | Algérie | ecq | est ce que |
| edr | Écroulé de rire | esk | est ce que | frr | frère |
| grv | grave | hbk | n7abek | hfdk | rabi yehafdek |
| hmd | elhamdoulilah | lsl | l'essentiel | jadr | j'adore |
| jdr | j'adore | jenémar | J'en ai marre | jms | Jamais |
| jsp | je ne sais pas | jtm | je t'aime | lol | rire à gorge déployée |
| lzm | lazem | mdr | Mort de rire | mr6 | merci |
| mrc | merci | msg | message | mzl | mazel |
| nbr | nombre | nhr | nhar | nn | non |
| omg | oh mon dieu | oqp | Occupé | pcq | parce que |
| pdq | pas de quoi | plz | please | prb | problème |
| prd | pardon | prq | pourquoi | prv | privé |
| Prx | prix | psq | parce que | PTDR | pété de rire |
| qlq | quelque | qoi | quoi | qqn | Quelqu'un |
| qsq | qu'est ce que | qst | question | r1 | rien |
| rdv | rendez-vous | ri1 | rien | sis | sœur |
| slm | salem | slt | salut | srt | surtout |
| srx | sérieux | stp | s'il te plaît | Svp | s'il vous plaît |
| Sys | système | tds | tout de suite | tel | téléphone |
| thx | thanks | Tjr | toujours | tkt | t'inquiète |
| tlf | téléphone | tlm | Tout le monde | tlv | television |
| ts | tous | v1 | viens | vazi | vas-y |
| wlh | wellah | wsh | wech | yhb | y7ab |

TABLE B.1 – Abbreviation list