

# Analysis Of Satisfaction Of Banque Populaire Customers Through Their Tweets

Youssef CHOUNI, Mohammed ERRITALI, Youssef OUADID

**Abstract:** The Social networks are an excellent source of information, and extraction of opinion. Nowadays, the most of internet users are using these platforms in order to share their sentiments and opinions about the products or services. The exploitation of these opinions is fruitfully. In this work, we expose the problem of sentiment analysis in social networks by showing the multiple experiments made in this context on Tweets using the two important approaches of this domain, namely, the Lexicon-Based Approach and the Machine Learning Approach. Also we introduce an original approach which incorporates the semantics in the second approach using the WorldNet lexical database.

**Index Terms:** Social networks, Twitter, Sentiment analysis(SA), WorldNet, language R, Machine learning, SVM, Naïve Bayes(NB).

## 1. INTRODUCTION

Since the creation of the Internet and until recently, websites have been platforms providing read-only information to most users. Following the advent of Web 2.0 and especially the massive use of social media, such as Twitter or Facebook, the number of users has increased considerably and their roles have evolved into that of providing information on the web. Social media has therefore provided a two-way information channel, the first creating a bridge between users and site owners, and the second between the users themselves. In the same trend, the banking sector has widely adopted social media as a means of communication to stay in touch with their customers and disseminate information. Traditionally, customers only had access to information when they were in front of a sales agent, or even in specialist journals written by experts. Currently, with the use of new technological means such as smartphones directly connected to social media, customers can search for information on the bank, participate in the debate on a related subject, appreciate a service or product from a bank organization. These interactions create a large volume of potentially usable information. The client is much more informed, has more and more power in the decision-making about the services which will be granted to him and can therefore choose which organization has, according to him, the best potential to meet his needs and administer the desired banking services. It is therefore important for these organizations to know how to collect data from social media and especially how to analyze this data in order to better understand their social ecosystem, and detect opinions and trends that would allow them to control their communication and improve the image that their customers have in relation to the services offered in order to know the satisfaction of their customers.

## 2 LITERATURE SURVEY

In recent years, several researches are carried out on the analysis of the feelings of tweeters. each method uses a specific approach either because of treated topics or the language of tweets or sometimes the difficulty of preparing the

- Youssef CHOUNI is currently Phd in TIAD Laboratory , FST Beni Mellal, Sultan Moulay Slimane University. E-mail: chouniy@yahoo.fr
- Mohammed ERRITALI is currently associate professor in TIAD Laboratory , FST Beni Mellal, Sultan Moulay Slimane University. E-mail: [m.erritali@usms.ma](mailto:m.erritali@usms.ma)
- Youssef OUADID is referent trainer in web development in Youcode, Youssoufia Morocco. E-mail: [yo.ouadid@gmail.com](mailto:yo.ouadid@gmail.com)

corpus. for Eugenio Martínez-Cámara & all, the best combination for the classification of the polarity of the tweets is to divide the tweets into unigrams, apply a stemmer to the unigrams, weight them with the TF measurement and use SMV as classification algorithm [1]. Also Richa Mathur & all have provided an analysis of the tweets related to Bitcoin to predict the price change in the near future; after the tests they have carried out with the 3 algorithms SVM, DT and NB they have deduced that SVM gives more accurate results than the others, as well as removing unwanted words in the pre-treatment phase, the construction time of the model is reduced and also provides more accurate results [2]. but this deletion is always good and reliable where it can sometimes change the meaning of sentences?

**TABLE 1**  
**USES LETTERS AS TOOLS OF NEGATION IN THE MOROCCAN DIALECT.**

Arabic	Moroccan dialect	English
لم يعجبني	معجبنيش	I Dislike
لم يفرحني	مفرحنيش	He didn't make me happy
غير فظيح	متخلعش	Not terrible

Moreover, Abdeljalil EL ABDOULI and all treated the topic of tweets analysis with another way. they are focused on an automatic collection method of a corpus that can be used to form a multilingual sentiment classifier so that it can classify tweets into positive and negative using the latent Dirichlet allocation (LDA) and classify him as Naive Bayes. but in this work there is no comparison with other algorithms to know the efficiency of choice, on the other hand one should not ignore the tools of negation especially the Moroccan dialect uses letters as tools of negation (Table1) as well as there is not a corpus that includes Moroccan words. [3] By the way Arun Manivannan and all, has chosen a hybrid classifier that consists in incorporating classifier emoticon, SentiWord classifier, and multinomial Naive Bayes classifier. to distinguish positive, negative and neutral tweets after a passage of two processing steps. Even if this method has worked well but one must ask the question of expression time, even if they did not mention it in the article but generally hybrid methods take longer. [4] In addition, Karthika K. shows that the Neural networks MLP and Random Forest are the best

classifiers with an accuracy of 94%; to analyze the feelings of sarcastic tweets because they have more flips of polarity, while the difficulty of determining a sarcastic Tweet without hash tag. however, in dialectic languages you cannot know whether a sentence is sarcastic or not, especially that each person expresses his opinion with a specific way, which makes the approach of detection of sarcasm a very heavy task. [5]

### 3 APPROACHES TO SENTIMENT ANALYSIS

Automatic approaches to sentiment analysis can be divided into two categories.

The literature review generally includes two approaches to the analysis of feelings:

- Lexicon based approach uses a predefined lexicon;
- The Machine Learning based approach uses the subjective classification of texts from a broad set of data.

#### 3.1 The lexicon-based approach to the analysis of feelings

Among the most well-known approaches in the literature is the lexicon-based approach to the analysis of feelings. Using two categories of predefined and weighted words, called lexicon or dictionary, one can identify the polarity of a text. Based on the dictionary, which is subdivided into two categories, one contains the positive and the other contains the negative. [6][7] It identifies all positive or negative words in a text. The dictionary consists of a small set of opinion words subdivided into two categories. One category contains words with more positive terminology, while the other category includes words with more negative feelings. At the end by calculating the sum of the words to have an overall assessment of the feeling in the text.

The algorithm used in the lexicon-based approach can be expressed as follows:

1. Initialize total feeling score  $s = 0$
2. For each text token, verify the presence in the lexicon:
  - a. If the token is present:
    - i. If the token is positive, then:  $s = s + w$
    - ii. If the token is negative, then:  $s = s - w$
3. Look at the total feeling of the text
  - i. If  $s >$  at the threshold, then the feeling of the text is positive
  - ii. If  $s <$  at the threshold, then the feeling of the text is negative.

#### 3.2 The learning machine approach

this approach uses classification algorithms; we can build a model from a database that contains transactions already verified to predict other data in the future. There are several algorithms in the literature, which is divided into two large families (supervised and unsupervised learning). And sometimes we can bring together two or more methods to create our own model, and we do tests to find out the best model that varies depending on the field of application and the data processed. [8]

### 5 PROPOSED WORK

Before starting the analysis of feelings, one must first have a clear vision on the documents, sentences or words dealt with. phrases in general are either objective or subjective, we focus on subjective phrases to determine positive, negative or neutral polarity. But at this stage we have to ask the question: at what level should we do our analysis? is it the level of the document (Message level or Document level), or the level of

the sentence (Sentence level) where one goes down to the lowest level, the level of aspects (Entity and aspect level). [9] And to answer this question, we need to know the interest of the study as well as the type of data dealt with. for example, if we want to know customer satisfaction for a product, where the views of specialists on a subject we can analyze the entire database to conclude the general polarity; Furthermore, if we want a mapping of customers interested or not in a product to plan actions for specific categories or regions, we must analyze the phrases and aspects. on the other hand, there are subjects that require a fine analysis of feelings, to see the degree of customer satisfaction (very positive, Positive, Neutral, Negative, very negative). Also there are programs that can mention whether feeling is a feeling of anger, sadness, happiness, or love.

#### 4.1 DATA COLLECTION

In order to collect customers' opinions on the popular bank in Morocco, we have two choices: either with forms (question, answer) or with data collection on social media (Facebook, tweeter, etc.). For the first choice which is a classic choice we find that most customers do not answer the forms and sometimes they ignore the emails which contain several questions, also sometimes the form sending period does not give reliable advice from the services of the bank or the bank in general. On the other hand, the use of social media by users-customers is fundamental, as it allows them to share opinions and search for information on various topics. On various social media platforms, client users can express feelings and participate in discussions. Some users share information on bank-related topics, such as their experiences with agencies. They can thus appreciate a service, recommend a product. They may also complain about specific aspects of the institution, such as good governance, quality of service, operation in the IT department). These data are valuable after they are extracted, processed and analyzed. The company Twitter to put at our disposal API (programming interface) that allow to access by program the messages and to carry out statistical studies and data mining. You just need to have a developer account to access this API. You must receive 4 keys from Twitter: Consumer Key, Consumer Secret, Access Token, Access Secret. To start collecting tweets, we have the choice between two languages R and Python which are open source programming languages with a large community. As well as new tools are permanently added to their respective catalog. By comparing the two languages we find that the learning curve of python is more complicated than that of R, the types of matrices and the basic graphics are integrated in the base language to the language R on the other hand with python one must know Num Py, Pandas and matplotlib ... etc. Python libraries can be difficult to configure, while most R packages work immediately and easily found through CRAN.

We will then work under R using the "twitter" package to collect the tweets.

starting with the creation of the connection with Twitter with the command:

```
setup_twitter_oauth(consumer_key, consumer_secret,
                    access_token, access_secret)
```

The searchtwitter() function allows you to load tweets online. Several parameters are available. In our example below, we specify the keyword that will select the tweets contains @BP\_Maroc (the Moroccan popular bank), we limit the

number of messages extracted to n = 1000, we are interested in documents in French (Lang).

```
tweets_BP <- searchTwitter("@BP_Maroc", n = 1000, lang = "fr", resultType = "mixed")
```

Tweets processing is even easier when we go through a data frame structure via the twlist2df() command.

```
tweets_df <- twList2DF(tweets_BP)
```

We can then, for example, store these tweets in a CSV file:

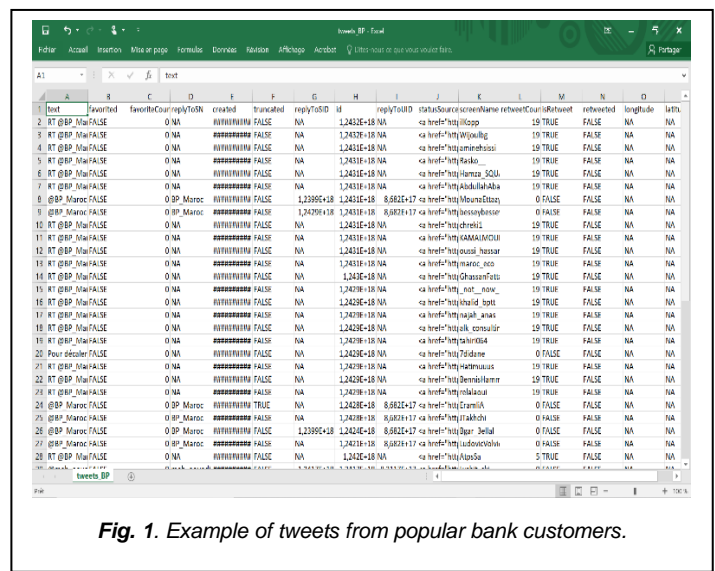


Fig. 1. Example of tweets from popular bank customers.

Each recorded line contains 16 parameters, for example the tweet identifier(id), the username to which it has responded(replyToSN), the tweet source(statusSource), the creation date(created)...etc. in our case we will look at the text itself of the tweet. example of tweet before pretreatment:

"@BP\_Maroc pourquoi sur votre application ou site mobile chabbinet on ne peut pas faire un virement qui dépasse 10000dh!!! C'est honteux"

**4.2 Data cleaning**

**a) Elimination of duplicates**

We eliminate repetition in messages  

```
tweets_df_U <- tweets_df[!duplicated(tweets_df$text),]
```

 we will store the messages (Tweet text) that we collect in a specific vector  

```
text_tweets <- tweets_df_U$text
```

**b) Removal of line breaks and web references**

Several elements that we will not exploit can disrupt. We remove them.

- removal of line break \n  

```
tweets_Clean <- gsub("\n", "", text_tweets)
```
- removal of Url  

```
tweets_Clean <- gsub("http\\S+\\s*", "", tweets_Clean)
```
- removal of excess spaces  

```
tweets_Clean <- gsub("\\s+", "", tweets_Clean)
```
- removal of the "\n"  

```
tweets_Clean <- gsub("[\n]", "", tweets_Clean)
```
- removal of spaces at the end of text  

```
tweets_Clean <- gsub("\\s*$", "", tweets_Clean)
```
- case harmonization - put everything in lowercase  

```
tweets_Clean <- tolower(tweets_Clean)
```
- removal of accents  

```
tweets_Clean <- gsub("['éèê]", "e", tweets_Clean)
```
- ```
tweets_Clean <- gsub("['ââ]", "a", tweets_Clean)
```

```
tweets_Clean <- gsub("[ùû]", "u", tweets_Clean)
```

removal of retweet indicator  

```
tweets_Clean <- gsub("rt ", "", tweets_Clean)
```

 deletion of words that contains hashtag (#) the bank, or bank references like @BP\_Maroc  

```
tweets_Clean <- gsub("#*", "", tweets_Clean)
```

```
tweets_Clean <- gsub('@bp_maroc', "", tweets_Clean)
```

automatically delete from all punctuation characters, also a good part of the French words "basic", such as "le", "la", ...etc then all empty spaces  
 library(NLP)  
 library(tm)  

```
tweets_corpus <- VCorpus(VectorSource(tweets_Clean))
```

```
tweets_corpus <- tm_map(tweets_corpus, removePunctuation)
```

```
tweets_corpus <- tm_map(tweets_corpus, stripWhitespace)
```

**Example of tweet after preprocessing:**

" application site mobile chabbinet faire virement depasse 10000dh cest honteux"

So far we have managed to reduce the corpus to 47%. In the following table we show some examples of the tweets before and after the pretreatment.

In the rest of our work we thought of reducing the corpus by using one of the following two methods: lemmatization or stemming. The first (lemmatization) consists in grouping the declinations of a lemma within the same entry in our corpus.

| Before                                                                                                                                            | After                                                                                |
|---------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| @BP_Maroc Veuillez améliorer vos canaux digitaux d'abord et essayer d'être réactif aux réclamations de vos clients !                              | veuillez ameliorer canaux digitaux dabord essayer detre reactif reclamations clients |
| @BP_Maroc On veut de l'argent pour survivre<br>Pour décaler vos traites chez @BP_Maroc c'est par ici https://t.co/CCoKFC6AB                       | veut largent survivre<br>decaler traites essayer detre reactif reclamations clients  |
| @BP_Maroc j'ai essayé de payer des factures d'eau, je tombe sur ce message d'erreur, j'ai refait l'opération 3 fois...<br>https://t.co/EeikHZf8VZ | essaye payer factures deau tombe message derreur refait loperation                   |
| @BP_Maroc pourquoi sur votre application ou site mobile chabbinet on ne peut pas faire un virement qui dépasse 10000dh!!! C'est honteux           | application site mobile chabbinet faire virement depasse 10000dh cest honteux        |
| @BP_Maroc Et pour l'ouverture de compte?                                                                                                          | ouverture compte                                                                     |
| @BP_Maroc j'arrive ps a me connecter                                                                                                              | arrive connecter                                                                     |
| @meh_aouadi @BP_Maroc Je veux une réponse de la banque Chaabi.. pour des gens qui ont payé une assurance toute leur vie.                          | veux reponse banque chaabi gens paye assurance vie                                   |
| @Hatimuus @BpMaroc @BP_Maroc de prendre soin de vos employés en agence qui souffrent. Aucune rotation d'effectifs...<br>https://t.co/PzzeLE7SvQ   | prendre soin employes agence souffrent aucune rotation deffectifs...                 |
| @BP_Maroc S'ils vous plaît, on a besoin de plus d'informations concernant le dépôt des demandes pour reporter le pa...<br>https://t.co/jKuvvmPPK  | sils plaît besoin plus dinformations concernant dépôt demandes reporter              |
| @BP_Maroc Cela fait presque 3 mois que mon incident n° 26446 est ouvert sans mise à jour ni solution.                                             | presque mois incident ouvesans mise jour solution                                    |

(Reduce the different forms of the word plural, feminine, conjugation, etc.). The second (stemming) is to reduce a



word to its root (stem), but sometimes stemming can lead to erroneous clustering (e.g. marmite, marmaille → marm) After the use of lemmatization, it has been found that the corpus reduction has not achieved a great success since most of the tweets used abbreviated words and sometimes we find that there are keystrokes errors, with a complexity of the corrected in an automatic way (possible that will be our objective in the next work). for this we will move on to the second classification step to analyze the feelings of tweets.

**4.3 Classification**

At this stage, we first examine the subjectivity of each text collected and formatted within our corpus. Only subjective texts are taken into account. Once we have all the subjective phrases, we have to determine what measures are taken and then evaluate them. For this study, the first is the feeling of each of its sentences in the database. Feelings analysis is done using one of the two Naive Bayes and SVM algorithms.

**a) Naive Bayes classifier**

This method consists in selecting the category that maximizes the retrospective probability of observing the text terms. For the learning phase we have a database which contains the words in French classified in 3 classes (positive, negative and neutral). NB is a simple Bayesian supervised classifier which assumes that all attributes are independent of each other, taking into account the context of the class [10]. Because of this assumption, the parameters of each attribute can be learned separately, and this greatly simplifies learning, especially when the number of attributes is large. Based on the NB hypothesis, probability of a document since its class can be calculated as the product of the probability of attribute values on all word attributes. In our case the bag of words that we created by the corpus contains several words, therefore the purpose of this part and calculate the probability of the words we did the test on the data set of test (contain the positive and negative words) to know the polarity of the words of the tweets.

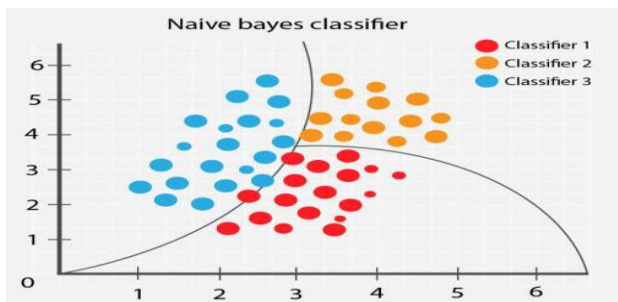


Fig. 2. Naive Bayes Classifier

**b) Support Vector Machine(SVM) classifier**

Support vector machines (SVM) This technique is a two-class classification method that tries to separate positive and negative examples in the set of examples. The method then looks for the hyperplane that separates the positive and negative examples, ensuring that the margin between the closest of the positive and the negative is maximum. This ensures a generalization of the principle because new examples may not be too similar to those used to find the hyperplane but may be located on one side or the other of the border. The interest of this method is the selection of support vectors that represent the discriminating vectors through which

the hyperplane is determined. The examples used in the hyperplane search are no longer useful and only these support vectors are used to classify a new case, which can be considered an advantage for this method. [2]

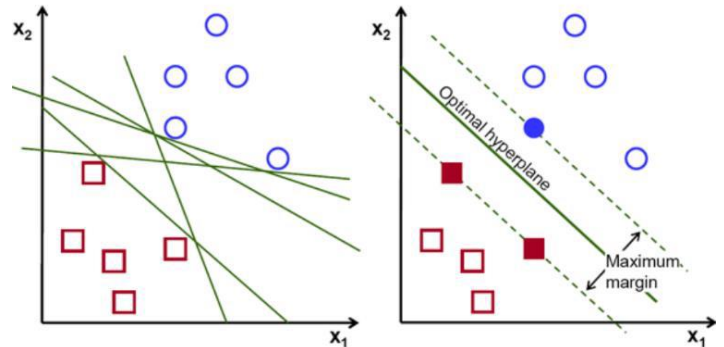


Fig. 3. Find the hyperplane linearly separates the classes

The NB classifier surpasses SVM with their precision in the learning phase. It can therefore be concluded that machine learning approaches can be one of the good alternatives for the classification of feelings. The following table shows the accuracy of each algorithm.

**TABLE 3**  
COMPARISON BETWEEN NAÏVE BAYES AND.

| Algorithm | Naive Bayes | SVM   |
|-----------|-------------|-------|
| Precision | 83,53       | 79,98 |

**4.4 Calculates the score**

We calculate two different scores for the polarity of the sentence. One score contains the addition of the score of the words found positive (noted score p), the other contains the addition of the score of the words found negative (noted score n). In our case, we calculate positive and negative scores by addition and create two separate dictionaries. Thus we are not interested in the polarity of the sentence but of entire corpus. So we will calculate the neutral terms to have a global view on the totality of the terms found.

Score [positive] = + score p  
 Score [negative] = + score n

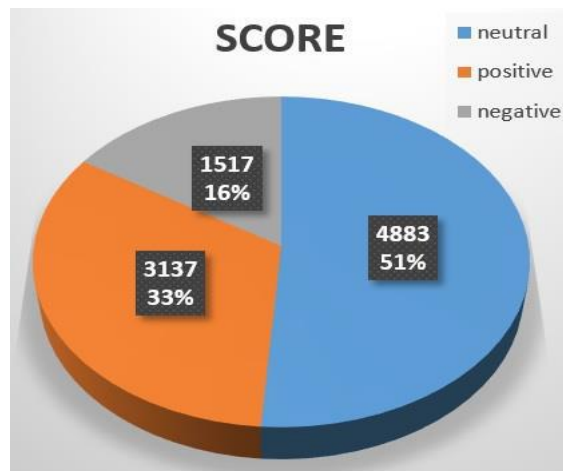


Fig. 4. Client satisfaction rates of the People's Bank

## 5. CONCLUSION AND FUTURE WORK

This paper presents a study of the feelings of tweets, including data collection, preprocessing, retrieval and classification. The pre-processing phase will take up much of this process and influence model success; most of the work deals with tweets in English or French, which makes work a little bit easier because of the existence of libraries that can help in the extraction and classification phases. On the other hand, in Morocco we find that most tweets are Arabic words written in French letters. In our case we focused on the classification of tweets using the naive classifier Bayes and SVM. We found that NB is doing better (83,53 precision). In our next work we will focus on the extraction phase to improve the process of creating the corpus of Arabic words or French words in dialect jargon.

## REFERENCES

- [1] Eugenio Martínez-Cámara & all "Polarity classification for Spanish Tweets using the COST corpus" *Journal of Information Science*, 2014, pp. 1-11 © The Author(s), DOI: 10.1177/0165551510000000.
- [2] Richa Mathur & all "Analyzing Sentiment of Twitter Data using Machine Learning Algorithm" *GADL Journal of Inventions in Computer Science and Communication Technology (JICSCT)* Volume-4, Issue-2 2018.
- [3] Abdeljalil EL ABDOULI & all "Sentiment Analysis of Moroccan Tweets using Naïve Bayes Algorithm" *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 15, No. 12, December 2017.
- [4] Arun Manivannan & all "Hybrid classifier for analyzing Twitter Sentiment"
- [5] Karthika K. "An efficient approach for sarcasm detection in tweets using polarity flip" *International Journal of Advance Research, Ideas and Innovations in Technology* ISSN: 2454-132X Impact factor: 4.295 (Volume 5, Issue 2) 2019.
- [6] Taboada & all "Lexicon-Based Methods for Sentiment Analysis" *Computational linguistics* 37, no. 2 (2011).
- [7] Amit Agarwal & all "Application of Lexicon Based Approach in Sentiment Analysis for short Tweets" *International Conference on Advances in Computing and Communication Engineering (ICACCE-2018)*.
- [8] Yassine Al-Amrani & all "Sentiment Analysis using supervised classification algorithms" *The 2nd international Conference on Big Data, Cloud and Applications* 2017.
- [9] Rahim Dehkharghani "Building Phrase Polarity Lexicons for Sentiment Analysis" *International Journal of Interactive Multimedia and Artificial Intelligence* 2018.
- [10] D. Davidov & all "Enhanced sentiment learning using twitter hashtags and smileys", *Proceedings of Coling*, 2010