Sentiment Analysis of French Political Tweets: #MacronPrésident

Antoine Vanrysselberghe and Els Lefever
LT^3, Language and Translation Technology Team
Ghent University, Belgium
Groot-Brittanniëlaan 45, 9000 Ghent
Email: {antoine.vanrysselberghe,els.lefever}@ugent.be

***Abstract*—The perpetual democratization of the Internet has made web user opinions on a wide range of topics continuously grow in value. As a result, many approaches to automatically analyse this user generated data have emerged over the last two decades. Sentiment analysis, in particular, aims to detect the presence of positive or negative sentiment within text. In this pilot study, we implement sentiment analysis on 615 political French tweets that all relate to the current French president, Emmanuel Macron. The experimental results show a satisfying performance of the supervised machine learning approach given the moderate size of the corpus. At the same time, the results reveal that the unequal distribution of the sentiments within the corpus (66% negative sentiment labels) considerably impacts the performance of the system for the positive and neutral sentiment labels. This pilot study shows, however, that supervised machine learning is a viable way to detect the global opinion of the French citizens on their president.**

Keywords–sentiment analysis; French political tweets; social media analysis.

I. Introduction

For a long time, mankind has eagerly tried to understand how people feel and how to decipher hidden meanings behind any type of utterance. The advent of the web 2.0 has without any doubt helped us to get closer to these goals. In that respect, sentiment analysis refers to an automatic Natural Language Processing (NLP) method that aims to detect the presence of a sentiment or an emotion in text and to classify this text according to a certain polarity (positive, negative or neutral). In addition, this field of research incorporates a great deal of additional dimensions. Researchers have, for example, created systems that are able to detect subjectivity, the type of expressed emotion or even the intensity of the emotion in question [1].

According to Pang and Lee [2], automatic sentiment analysis starts to draw the attention in 2001. The many opportunities that this method brings to the table and the many doors that it could end up opening start to catch the interest of various domains such as the political, sociological, financial, governmental, publicity and marketing domains. As Boullier and Lohard [3] point out, sentiment analysis now forms “an evident resource for any marketing or communication team of a brand”. That being said, this steadily growing interest is not only linked to the development of sentiment analysis. The progress in the field of machine learning, the expansion of the web together with the rise of ‘Big Data’, as well as the ever increasing ability to vent your personal opinion on the internet have all played a considerable role towards that development [4].

Social media platforms like Facebook and Twitter have become increasingly popular in the NLP research community. In fact, this scientific interest can be regrouped under multiple primary reasons. First of all, the content on these networks is almost entirely user-generated, which also potentially means that any message that has been posted contains an opinion or some trace of subjectivity. Additionally, the user is free to post whatever he wants, what leads to a wide array of themes and subjects that, for the most part, are generally linked to current events. This can, for example, be seen via the ‘trending hashtag’ section on the Twitter homepage. Boullier and Lohard [3] observe that this cluster of opinions at a certain point in time can also be used to analyse the evolution of opinions on a certain subject, given that such analysis occurs on a regular basis. Secondly, the application programming interface of the vast majority of these social media platforms is openly accessible, which makes data collection very easy, as stated by Fang and Zhan [5]. Finally, the amount of people that generate content online is massive. As an example, Twitter and Facebook together boast more than 1500 million active users a day (100 million and 1400 million, respectively) [6]. Such social media platforms can therefore be seen as gigantic data mines that allow researchers, businesses as well as other individuals to collect data on a wide spectrum of different subjects. These parties have the opportunity to analyse the behaviour of shareholders within the stock market, the political tendencies of specific countries or even the development of new internet phenomena such as GIFs (Graphics Interchange Format) and memes [7]. In addition, specialised websites have even been created to gather the opinion of web users on a multitude of specific domains. Some examples are IMBD, Rotten Tomatoes, the video-gaming platform Steam or even Myanimelist for everything related to manga and Japanese animation. Multiple companies have created a core business out of annotating data for sentiment analysis. On the one hand, businesses have specialised themselves in data collection for their clients. On the other hand, companies, such as Indico, AYLIEN and Nexosis offer software that allow other companies to easily collect and analyse data themselves. Moreover, phone apps allow for an even easier data collection because of their Application Programming Interface (API). Furthermore, as Chevalier and Mayzlin [8] point out, customer reviews that are present on commercial websites influence the choice of future customers. A great deal of positive reviews on a certain product is therefore considered a major asset that encourages a purchase. Some online price comparison tools even present their users with an overview of positive and negative reviews for the requested product.

In the political field, which is also the main field this study focuses on, sentiment analysis can be employed to detect the public opinion, to recognize the sentiments expressed about a certain candidate, to predict the outcome of an election or even

to predict the coalitions that will be formed during or after the elections.
The remainder of the paper is organized as follows. Section II gives a brief overview of existing research and methodologies of sentiment analysis, while Section III describes the corpus that was compiled and manually annotated. Section IV describes our experimental setup and the different feature groups we used. In Section V, we report on the results of our sentiment analysis classifier and perform a detailed error analysis. Section VI concludes this paper.

II. Related research

Two main paradigms have been applied to conduct sentiment analysis [7]. The first method is the lexical-based approach, where sentiment analysis is performed by looking up word combinations and the sentiment that these combinations evoke within a sentiment lexicon. The major flaw of this approach, which is still widely applied within commercial sentiment analysis systems, lies in the fact that the list of words that are used do not include any kind of context. As a result, ambiguous words are especially difficult to map. An example of a well-performing lexicon-based approach is VADER (Valence Aware Dictionary and sEntiment Reasoner) [9], a rule-based model using lexicons, which are more sensitive to sentiment expressions in social media contexts. The second approach, which is also the one that has been put into practice during this research, is the corpus-based approach. The latter approach is for the most part integrated in machine learning, and aims to develop systems trained on (labeled) data that are able to independently attribute sentiment labels to new data, and this without any kind of human intervention. Various types of supervised approaches have been applied for sentiment analysis (a.o. Support Vector Machines, Nave Bayes, decision trees), incorporating a variety of lexical, syntactic and topical features [10]. More recently, deep learning approaches, which are inspired by the structure of the biological brain, have emerged as powerful machine learning techniques that have successfully been applied to sentiment analysis [11]. At the same time, researchers have started to investigate sentiment analysis at a more fine-grained level. Aspect-based Sentiment Analysis (ABSA) [12] [13], aims to extract (and summarize) the opinion people have on specific entities and on the aspects of said entities. It might, for instance, be the case that a user rating a new smartphone likes the battery, but thinks negatively of the screen. Although the general sentiment on the product can be positive or negative, the user might have different opinions on the different aspects of the product. This fine-grained ABSA is a challenging task, requiring automatic aspect extraction, aspect categorisation and sentiment analysis. Other important initiatives driving the sentiment analysis research are the multiple joined sentiment analysis projects that have emerged, such as SemEval (Semantic Evaluation) and DEFT (le Défi fouille de textes). These projects, that often take the form of conferences and workshops, focus among other things on hot topics within the field of sentiment analysis, such as for instance ABSA. Because of the nature of these joint projects, multiple research teams try to develop systems that aim to find the best solution for one of the many research questions, which results in some kind of friendly internal competition. All of this ultimately leads to an important amount of new research data and a comparative overview of all the systems which allows the community to unravel the methods that work best for a given task. Despite the fact that these projects have, for the most part, focused on English data, the spotlight has recently been turned on multiple foreign languages, such as Arabic, French, Dutch and even Chinese [12].
Conducting sentiment analysis on data collected from social media is bound to a couple of global difficulties that are generally not present in other data or document types. This is, for the most part, due to the fact that the core message of a tweet is often strengthened by elements that are not made up of words, such as images, GIFs and links to videos on other online platforms, for example. Furthermore, web users write, for the most part, in a rather informal way. Because of that, abbreviations that are typical of the web 2.0 surface, such as ‘wdym’ (what do you mean), ‘ic’ (I see), ‘imma’ (I will / I am about to), etc. While frequently used abbreviations can easily be transformed back to their original or more formal form, the web is also a dynamic place where codes, ways of writing and habits keep evolving. For the past few years, more and more abbreviations with a foreign origin keep finding their way into the French internet language. Words such as ‘afk’ (away from keyboard), ‘omg’ (oh my god) and ‘wtf’ (what the fuck) are starting to appear way more frequently on social media. Moreover, compared to official or professional documents, user generated data may contain spelling mistakes and wrong grammatical constructions. Frequently recurring mistakes are for example made against the indefinite articles ‘tout’ and ‘tous’ as well as mistakes against the conjugation of verbs that follow each other consecutively. When the present perfect is used, for example, the past participle is often conjugated as a simple infinitive (*j’ai manger ma pomme*). The typical social media symbols, namely hashtags, at signs, smileys and emojis, on the other hand, do not pose a real problem. As a matter of fact, hashtags and at signs can be used to determine the frequency with which they appear in the corpus. This can, among other things, be used to detect which ones often appear together. Moreover, Barbosa and Feng [14] have proven that, by exploiting the typical social media functionalities to create new features, which they call microblogging features, their sentiment analysis system performs better.
Sentiment analysis has also been applied under different angles within the political domain. Williams and Gulati [15], for example, have analyzed the impact of Facebook during the presidential elections in the United States in 2008. This study shows that the electoral support that a runner up for the White House receives on Facebook accurately reflects the success of his electoral campaign. A correlation even exists between the online electoral support of a candidate in a certain country and the final voting results. Tumasjan et al. [16] have analyzed political tweets during the federal elections in Germany in 2009. This research showed that Twitter is indeed used as a platform for political debate and that this particular social media even reflects the political mainstream that is present in the physical world. They also note that the political tweets are not only used to vent personal opinions but are also a way to share and engage with the political opinions of the other users. Additionally, a correlation between the total number of tweets (100.000) and the final results of the elections was also found. The political parties that often emerged together in the tweets reflected the coalitions that were present at that moment in time. Nonetheless, it is important to note that a small number of users (4% in that particular analysis) were responsible for 40% of the tweets within the compiled corpus.

That being said, the task of sentiment analysis is more complex than it looks like. Compared to a human being, a machine struggles to detect figures of speech, ironic or sarcastic data. As Cambria and White [17] point out, sentiment analysis now embodies a unique domain that focuses on semantics and existing sentiments, which is situated somewhere between natural language processing and the understanding of that natural language.
In this paper, we conduct a pilot study to discover the efficiency of supervised machine learning to implement sentiment analysis on French political tweets. Emmanuel Macron, the current French president is the main topic of this research. His election and his proactive policy have led to a huge amount of tweets that can be analysed to detect the global opinion of the French citizens on their president.

III. CORPUS

To test the viability of sentiment analysis on French political tweets, a corpus of 615 French political tweets relating to the French president Emmanuel Macron has been manually compiled and annotated. Furthermore, two versions of the corpus were created: the first corpus contains the tweets stripped from smileys and hashtags, unless the hashtags is an integral part of a sentence. The second corpus contains the same tweets including smileys and hashtags. These two versions will allow us to analyze the impact of smileys and hashtags on automatic sentiment analysis.
As we intend to sample a global overview of the French opinion on Emmanuel Macron, tweets that were not linked to the president in any way were purposely excluded. To discover what the French think about him on different aspects, tweets were collected for four different categories: (1) the global opinion on Macron (hashtag #Macron was used together with the keyword 'Macron', 135 tweets), (2) the opinion on one of his newest laws, the housing-tax reform (hashtag #taxed-habitation and keywords 'taxe', 'habitation' and 'Macron', 124 tweets), (3) the opinion on his appearance during an exclusive interview on France 2 (#Macronjt20HWE, 210 tweets) and (4) the opinion about Macron during the elections (hashtag #macronprésident, 146 tweets).
All tweets were labeled with a sentiment label (positive, negative, neutral) and with an indication of whether a trace of irony was present in the tweet or not. Figure 1 shows the distribution of the sentiment labels. The large amount of negative labels might be surprising given the fact that Macron has been elected president with 66.1% of the votes. It is good to keep in mind, though, that the French presidential election procedure consists of different rounds. In the first round, Macron's new political party *En Marche!* and Marine Le Pen's *Front National* pulled 24,01% and 21,3% of the votes, respectively. In the second round, a lot of French citizens voted for Macron to simply block Le Pen from ascending [18].
Some general observations could be made during the corpus compilation and annotation process. Similarly to [16], a specific part of the users was responsible for many different tweets on the given subject. As the opinion of a tweeter generally does not change in a few hours between the posts, it was decided to restrict the data inside the corpus to one tweet per user. As this present study attempts to extract the global opinion of the French citizens on their new president, including multiple positive, negative or neutral tweets of a single user would inevitably skew the representation of a global opinion. Similarly, various members of different political parties, both for and against Macron, have vented their opinion on Twitter. Logically, Macron's opponents reacted negatively towards him while his adherents were supporting him in a positive way. In spite of this, retaining only one tweet per user quickly balanced out this phenomenon.

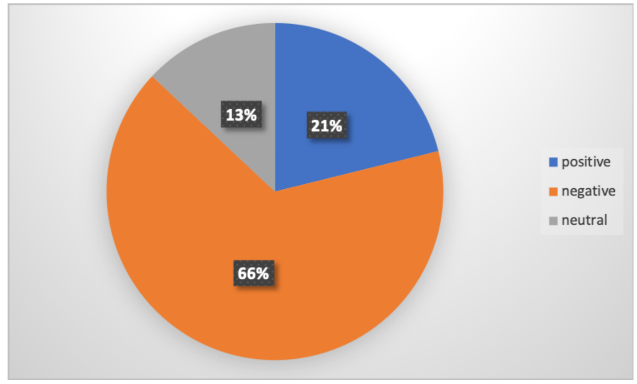

Figure 1. The distribution of the different sentiment labels in the corpus.

Moreover, around 5% of the tweets within the corpus are ironic. The vast majority of these ironic tweets also contain graphic elements, especially satires, to convey spot more efficiently. The system trained for this study, however, did not take into consideration these visual elements, which might have had a negative impact on its learning phase. Because the social media platform decided to double the character limit of a tweet from 140 to 280 characters two years ago, a very specific tweet pattern often recurred concerning the presidents interview on France 2. As the maximum amount of character increased and that the topic in question is an interview, users often inserted a specific quote of the president followed by their own opinion. Finally, when a refined search was used to further develop the corpus using the popular hashtags #macronprésident and #MacronJT20HWE, multiple spambots were encountered. These spambots abused the popularity of the hashtags to efficiently broadcast their off-topic messages. As the corpus was compiled manually, this type of data was successfully avoided.

IV. EXPERIMENTAL SETUP

We evaluated the feasibility of sentiment analysis of French political tweets by means of a supervised classification-based approach. We opted for the LIBSVM [19] with a linear kernel as our classification algorithm. In order to train and test the system, 10-fold cross-validation was implemented, where the data is divided into 10 equal folds, allowing 90% of the data to run as training and 10% of the data to run as test within each fold. The k Cross-validation method turns out to be especially efficient within the context of this study for two main reasons. First of all, the compiled corpus remains rather small and selecting one fixed test set could lead to less reliable results. Secondly, there is a notable imbalance of sentiments within the corpus (80 neutral tweets, 130 positive tweets, 405 negative tweets). To determine the efficiency of the system, the accuracy, precision, recall and F-scores (weighted average of precision and recall) were used as performance measures.
As a preprocessing step, all posts were tokenised using the

LeTs Preprocess Toolkit [20], and the following linguistic features were constructed:

1) **Token and character n-grams**. As a great deal of comparative studies have already demonstrated, basic linguistic features are simple yet very efficient [21]. Token-n grams (varying from 1 to 3 words) as well as character n-grams (ranging between 3-4 characters) were included for the experiments.
2) **Flooding**. Textual flooding happens when a user writes the same characters or words over and over again or when he abuses punctuation marks. This can also be used as a feature to detect a certain sentiment in a text. An unsatisfied user can, for example, repeat negatively connoted words to vent his opinion as in *"Les hommes politiques ne tiennent jamais leurs promesses, j'en ai marre marre marre marre marre marre !!!!!!!!" (English: Politicians never keep their promises, I've had enough!)*. Two features based on textual flooding were extracted: token flooding and punctuation flooding.
3) **Capitalisation**. Writing complete words or even sentences in capitals can refer to a specific sentiment. In the following example, the capitals are used to further convey the dissatisfaction of the user: *"LES MENSONGES DE MACRON S'ACCUMULENT ..." (English: "MACRON'S LIES ARE ACCUMULATING...)*.
4) **The NRC lexicon**. To capture sentiment words, the French part of the NRC lexicon [21] was included in the pipeline. This lexicon, developed as a shared project, proposes a list of more than 25.000 manually annotated English terms. Each term is also linked to a certain emotion as well as to a sentiment (positive or negative). This list has been automatically translated via Google Translate into more than 100 languages, including French. The lexicon is used for the look-up of all tokens in the tweet and stores the number of retrieved positive, negative and neutral words, as well as the polarity sum as features in the pipeline.

V. Results and Analysis

Given the moderate size of the corpus (615 instances) and the unbalanced representation of each sentiment (with the negative sentiment being the most prominent), the system achieves satisfying classification results. Table I shows the experimental results per fold for the corpus where smileys and hashtags have been removed, and the difference in performance (between brackets) with the corpus including smileys and hashtags. The value between brackets was calculated by subtracting the results of the second corpus from the first one (results corpus 1 - results corpus 2). Consequently, positive values indicate that the results of the first corpus were higher whereas negative values indicate that the results of the second corpus were higher.

Overall, the system achieves convincing accuracy results, with an average accuracy of around 70%. The accuracy of the corpus without smileys and hashtags averages 68.71% compared to an average of 69.19% for the corpus including these elements. While the smiley-hashtag corpus only achieves better results twice (in the first and tenth fold), the differences amount to 5% and 4.48%, which ultimately favors it.

A first striking observation concerns the zero values, which appear for two sentiments, viz. the positive and neutral sentiments. Because the positive and neutral sentiments only represent a third of the total corpus, two possible reasons can explain these zero values. Firstly, as the system had access to less positive and neutral training data, it seems to have had a harder time identifying these two sentiments. Secondly, the zero values for the positive sentiments only appear in the first and third folds, which contained very few positive tweets (five positive tweets and one single positive tweet in the first and third fold, respectively). This offers the system an extremely small error margin. The zero values for the neutral sentiment can be explained the same way. From the 300 tweets that these five folds contain, only 24 tweets contain a neutral sentiment. Moreover, the neutral sentiment remains the least represented sentiment in the corpus. Two actions could be taken to circumvent this problem. The first option would be to equally allocate the amount of positive, negative and neutral tweets in each fold. The second option would obviously be to expand the corpus to collect more positive and neutral tweets. This would allow the system to be better prepared for the detection of these two (underrepresented) sentiments.

It is clear that the system that was trained for this pilot study achieves much better results when it comes to the detection of negative sentiments. The third fold even contains results exceeding 90%. Furthermore, this fold only contains two non-negative tweets, which explains this almost flawless results. Similarly, folds 4 and 7 propose above average results. Yet, the precision of folds 1, 8 and 10 are rather mediocre (66.67%; 58.70% and 59.30%, respectively). The system often makes the same type of mistake inside these folds: it has a hard time attributing the correct label to tweets that are formulated as questions as well as to positive tweets that contain negatively connoted words. In the latter case, the system falsely labels the tweet as containing a negative sentiment, as is the case for the following example: *L'ingratitude est le signe des leaders implacables. Aurait-on enfin un bon président? (English: Ingratitude is the sign of relentless leaders. Would we finally have a good president?)*.

In fold 8, the system often failed to predict a positive label. As a result, we can observe a mediocre precision of the negative sentiment (58.70%), which correlates with the mediocre recall of the positive sentiment (40.63%). Once again, these errors appear most of the time when a positive tweet contains negatively tainted words: *Non, pas d'accord. Il a été pertinent. Comme d'habitude (English: No, I disagree. It was relevant. As usual.)*.

Finally, the system also struggled to analyze tweets containing abbreviations and figurative language, as in the following two examples: *L'itw de #MacronJT20HWE un grand coup de com sur un fond de politique d'austrit (English: The #MacronJT20HWE itw(interview), a publicity stunt hidden behind a political background focusing on austerity)* and *" Un âne aurait l'étiquette En Marche, il aurait été élu." (English: Should a donkey wear an 'En Marche' sticker, it might very well be elected)*.

Despite the system achieving very high results for the prediction of the negative sentiment throughout the analysis, this is not the case for the positive and neutral sentiment. Apart from the zero values that have already been discussed, the majority of the results for the positive sentiment vary between 25% and 50% with some performance peaks in the last three

 ISBN: 978-1-61208-725-2

TABLE I. ACCURACY, PRECISION, RECALL AND F-SCORES FOR THE POSITIVE (POS), NEGATIVE (NEG) AND NEUTRAL (NEU) SENTIMENT LABELS (IN PERCENTAGE) FOR THE SYSTEM WITHOUT SMILEYS AND HASHTAGS. THE VALUES BETWEEN BRACKETS REFLECT THE DIFFERENCE IN PERFORMANCE WITH THE SYSTEM INCLUDING SMILEYS AND HASHTAGS.

Fold	Accuracy	P_POS	R_POS	F1_POS	P_NEG	R_NEG	F1_NEG	P_NEU	R_NEU	F1_NEU
1	61.67 (-5.0)	0.00	0.00	0.00	64.71 (-2.0)	86.84 (-7.9)	74.15 (-4.1)	44.45 (-22.2)	21.05	28.57 (-3.4)
2	63.93 (1.6)	25.00 (5.0)	50.00	33.33 (4.8)	70.83 (0.6)	85.00 (4.5)	77.27 (2,3)	44.44	21.05 (-1.2)	28.57 (-1.1)
3	88.52	0.00	0.00	0.00	96.43 (-0.1)	91.53 (-3.0)	93.91	0.00	0.00	0.00
4	73.77 (1.6)	28.57 (3.6)	25.00	26.67 (1.7)	79.63 (0.4)	97.73 (2.3)	87.76 (1.2)	0.00	0.00	0.00
5	63.93	25.00 (1.9)	30.00	27.27 (1.2)	75.00 (-1.6)	81.82	78.26 (0.9)	0.00	0.00	0.00
6	70.49 (1.6)	52.94 (2.9)	64.29 (14.3)	58.06 (8.1)	82.5 (1.6)	76.74 (-2.3)	79.52 (-0.5)	25.00 (5.00)	25.00	25.00 (2.8)
7	70.49	53.85 (-4.5)	43.75	48.28 (-1.7)	76.60 (3.1)	92.31	83.72 (1.9)	0.00	0.00	0.00
8	65.57	92.86	40.63	56.52 (-1.3)	57.45	96.43 (-1.0)	72.00	0.00	0.00	0.00
9	70.49	73.33 (6.7)	57.89 (-5.3)	64.71 (-0.2)	69.77 (-2.7)	85.71 (2.9)	76.92 (-0.4)	66.67	28.58	40.01
10	58.21 (-4.5)	71.43 (-14.3)	20.00 (-4.0)	31.25 (-6.3)	55.93 (-3.4)	94.29 (-5.7)	70.21 (-4.3)	100.00	14.29	25.01

folds, especially fold 8 with an accuracy of 92.86%. As was already mentioned, these low results can be attributed to two deciding factors: the small corpus size and the unbalanced representation of the sentiments within the folds. The five first folds, which also contain the lowest results, only contain 25 positive tweets in total. The last folds, however, provide much higher results as many more positive tweets are present within the data. These results show that when the system determines that a tweet is positive, the given tweet is in fact positive. Yet, all the folds contain low recall scores.

A next point of analysis concerns the impact of smileys and hashtags on the sentiment analysis. Globally, the differences in performance between the first corpus (without smileys or hashtags) and the second corpus (with smileys and hashtags) are minimal. When looking at it in more detail, two major differences can be noted. First of all, the system trained on the first corpus is better at predicting positive sentiments. Secondly, the second system performs better when it comes to detecting negative sentiments.

The performance gaps are especially noticeable for the positive sentiment, and more specifically for fold 6 (a difference of 14.29% for the recall and 8.06% for the F1-score), fold 9 (a difference of 6.66% for the precision and 5.27% for the recall) as well as for fold 10 (a difference of 14.28% for the precision and 6.25% for the F-score). The performance gaps for the negative sentiment, on the other hand, are more discreet and hover around 2% to 3%. The recall results, however, fluctuate noticeably on three occasions. In fact, a 7.90% difference can be observed in the first fold, a 4.51% disparity in the second fold as well as a divergence of 5.71% in the last fold. The results for the negative sentiment, in comparison, barely differ. Despite 1%-3% differences appear here and there, the major gap resides in the precision of the first fold. In this case, a 22.22% disparity is present in favor of the second corpus.

This enhanced performance for the detection of the negative sentiment can be attributed to two determining factors. Firstly, the corpus contains many negative tweets, which entails that more negatively tainted hashtags and smileys have been added in comparison to the other sentiments within the second corpus. Furthermore, a more detailed comparative analysis has shown that users are more inclined to use emojis to express a so-called negative sentiment. The most common smileys are: the angry emojis, the crying emojis as well as the crying of laughter emojis (to express some kind of disbelief). The latter is often accompanied by a touch of irony.

As was already mentioned, figures of speech such as irony can pose a problem for machine learning systems that have not specifically been configured for it. In total, the corpus for this pilot study contains 30 ironic tweets (e.g., *Vive le roi ! (English: Long live the king!)*). In total, the system attributed a wrong label to an ironic tweet six times out of 30. In other words, the system attains an accuracy of 80% on ironic tweets. Three of the six mistakes were made when labeling a neutral tweet. In the following example, the system has attributed a neutral label to a negative tweet: *#Macron souhaiterait remplacer la taxe d'habitation par un nouvel impot qui serait plus juste mais comparable à la taxe dhabitation. Ce mec est un génie. (English: #Macron would like to replace the housing tax with a new tax that would be fairer yet comparable to the housing tax. This guy is a genius.)*. The first sentence indeed conveys a neutral sentiment at first glance. The second sentence, on the other hand, arguably conveys a negative sentiment that mocks the French president (*Ce mec est un génie*). Another example concerns a positive label that has been wrongly attributed to a negative tweet: *Ne dérangez pas Macron cette semaine , il finit de libérer la Syrie avec ses petits bras et il règle la faim dans le monde courant 2018 (English: Leave Macron alone this week, once he has finished freeing Syria with his little arms, he will have solved world hunger in the course of 2018)*. In this example, the irony clearly misleads the system. After having identified positive words such as 'libérer', the system labeled this tweet as being positive. Yet, 'ses petits bras' and mentioning the impossible task in such a short time-lapse clearly point towards mockery and discontent.

Another interesting example is the following one: *Quelle ébouriffante nouveauté. On avait pas vu a depuis...Guy Mollet en juin 1956 (English: What a mind-blowing innovation. We hadn't seen that since... Guy Mollet in June 1956)*. This example is quite interesting as it demonstrates how difficult cultural references are for automated systems. This specific reference, pointing towards Guy Mollets interview in June 1956, was one of the first interviews where a journalist was invited inside the president's office. This was something exceptional at the time. Similarly, Macrons interview in December 2017 was also held in the president's office. Yet, this time, this was seen as something old fashioned. As a result, this tweet hides a negative sentiment that the system was unable to infer.

VI. Conclusion

This paper presents a classification-based approach to sentiment analysis on French political tweets. To this end, a corpus of tweets concerning the current French president Emmanuel Macron has been collected and manually annotated. The experimental results and analysis show that the system developed for this pilot study achieves fairly good results given the limited corpus size, the amount of features used as well as the imbalance of the three main sentiments present in the data. Globally, the system achieves an average accuracy of 70% for both corpora. While the system does achieve high results for the detection of the negative sentiment, various improvements can be made to enhance the performance for the positive and neutral sentiment. First of all, much more data could potentially be added to the corpus as the 'Emmanuel Macron' topic generates a constant flux of new data. A larger corpus would also translate into much more training data concerning the positive and neutral sentiment. Secondly, ensuring an equal representation of each sentiment in each fold could potentially eliminate zero values and provide more streamlined results across all sentiments. Thirdly, incorporating more advanced linguistic features, such as common-sense knowledge [22], could help to improve the sentiment analysis accuracy.

A manual error analysis has revealed multiple causes for the wrong prediction of positive and neutral tweets. The presence of negatively connotated words inside a sentence, tweets formulated as a question, abbreviations, cultural references as well as figurative language all seemed to complicate the correct prediction of sentiments. Although irony usually poses a major problem in sentiment analysis, the system achieved convincing results with an accuracy of 80%.

Despite the unbalanced distribution of the sentiment labels in the corpus and the resulting classification problems, the lack of positive and neutral data on the current French president reflects a reality that cannot be ignored. Macron's election was a rather surprising event that occurred without a real majority vote. According to multiple sources, his election can partially be attributed to a wish to block the Front National (FN). In creating a small corpus on a specific subject, the distribution of the sentiments present in the collected data closely reflect the general opinion on the subject. In present case, a negative opinion of 70% correlates to the most recent opinion polls on Macron at the time of the study. The small corpus size can therefore be considered as a representative sample of the online public opinion. This, together with the high results for the negative tweets in comparison to the positive and neutral ones, reflect in present case the discontent and distrust of the French population towards Macron.

In conclusion, using a larger corpus for this study with a better balance of the sentiments would most certainly lead to better results, especially for the positive and negative sentiments. Overall, we can conclude that a support vector machine with a linear kernel is a viable way to perform sentiment analysis on French political tweets.

References

[1] F. Benamara, M. Taboada, and Y. Mathieu, "Evaluative language beyond bags of words: Linguistic insights and computational applications," Computational Linguistics, vol. 43, 2017, pp. 201–264.

[2] B. Pang and L. Lee, "Opinion mining and sentiment analysis. Foundations and Trends," Information Retrieval, vol. 2(1-2), 2008, pp. 1–135.

[3] D. Boullier and A. Lohard, Opinion mining and Sentiment analysis. OpenEdition Press, 2012.

[4] C. Van Hee, Sentiment Analysis for political tweets: a corpus study (original title: L'analyse des sentiments appliquée sur des tweets politiques: une étude de corpus). Ghent University, 2013, phD.

[5] X. Fang and J. Zhan, "Sentiment analysis using product review data," Journal of Big Data, vol. 2(1), 2015, pp. 1–14.

[6] "Twitter by the Numbers: Stats, Demographics Fun Facts," 2018, URL: https://www.omnicoreagency.com/twitter-statistics/ [retrieved: Juin, 2019].

[7] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha, "Comparing and combining sentiment analysis methods," in Proceedings of the first ACM conference on Online social networks. ACM, 2013, pp. 27–38.

[8] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," Journal of marketing research, vol. 43(3), 2006, pp. 345–354.

[9] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in Proceedings of the 8th AAAI conference on weblogs and social media (ICWSM), 2014, pp. 216–225.

[10] M. Joshi, P. Prajapati, A. Shaikh, and V. Vala, "A survey on sentiment analysis," International Journal of Computer Applications, vol. 163(6), 2017, pp. 34–38.

[11] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018. [Online]. Available: https://doi.org/10.1002/widm.1253

[12] M. Pontiki et al., "SemEval-2016 task 5: Aspect based sentiment analysis," in Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), 2016, pp. 19–30.

[13] O. De Clercq, E. Lefever, G. Jacobs, T. Carpels, and V. Hoste, "Towards an integrated pipeline for aspect-based sentiment analysis in various domains," in Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2017, pp. 136–142.

[14] L. Barbosa and J. Feng, "Robust Sentiment Detection on Twitter from Biased and Noisy Data," in Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010), 2010, pp. 36–44.

[15] C. B. Williams and G. J. Gulati, "Facebook grows up: An empirical assessment of its role in the 2008 congressional elections," in Annual Meeting of the Midwest Political Science Association, 2009, pp. 2–5.

[16] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment," in Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, 2010, pp. 178–185.

[17] E. Cambria and B. White, "Jumping NLP curves: A review of natural language processing research," IEEE Computational Intelligence Magazine, vol. 9(2), 2014, pp. 48–57.

[18] "Presidential Election: vote white or Macron, LR excludes abstention regarding the FN (original title: Présidentielle: voter blanc ou Macron, LR exclut l'abstention face au FN)," 2017, URL: https://www.lexpress.fr [retrieved: Juin, 2019].

[19] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2(3), 2011, pp. 1–27.

[20] M. Van de Kauter et al., "LeTs Preprocess: the Multilingual LT3 Linguistic Prepro- cessing Toolkit," Computational Linguistics in the Netherlands Journal, 2013, pp. 103–120.

[21] S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry, "Semeval-2016 task 6: Detecting stance in tweets," in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016, pp. 31–41.

[22] E. Cambria, D. Olsher, and D. Rajagopal, "SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis," Twenty-eighth AAAI conference on artificial intelligence, 2014.