Sentiment Analysis on Post conflict in Colombia: A Text Mining Approach

Víctor Daniel Gil Vera[1*], Lina María Montoya Suárez[2], Isabel Cristina Puerta Lopera[1]

[1] Universidad Católica Luis Amigó
Tv. 51a #67B-90, Medellín, Colombia

[1] Corporación Universitaria Remington / Universidad Católica Luis Amigó
Calle 51 No. 51-27 / Tv. 51a #67B-90, Medellín, Colombia

[*]*Corresponding author's email: victor.gilve [AT] amigo.edu.co*

ABSTRACT—*The post conflict in Colombia is the period of time that began with the agreement signed in 2016 between the government and the Revolutionary Armed Forces of Colombia (FARC-EP) where it is agreed the cessation of the armed conflict and the establishment of a stable and lasting peace. The objective of this paper is to perform a tweets analysis of feelings and opinions from the social network Twitter related to the Colombian post conflict, which has become a widely debated topic in the world. 250 tweets of Colombians and 250 tweets of Foreigners were collected. The feelings of Foreigners were considered to analyze the public perception that other people have different to Colombians. A comparative analysis was developed; the results of the Colombians were significantly different from Foreigners. The tweets of Foreigners have more positive feelings compared to the Colombians tweets. In comparison with the proportion of tweets with positive feelings of Foreigners, 60%, the percentage of Colombians tweets is only 20%.*

Keywords— Colombia, Peace, Post conflict, Sentiment analysis, Text mining.

---

1. INTRODUCTION

The internal conflict in Colombia is a problem that had a duration of nearly half a century, in its majority was caused by criminal groups outside the law who rebelled against the government. This issue generated a large number of deaths, displacement, kidnapping, torture, rape, etc., problems that turned this country as one of the most dangerous in Latin America and the world. In the Introduction section, present clearly and briefly the problem investigated, with relevant references. The main results should be enunciated.

The objective of this paper is to perform a sentiment analysis of Colombians and Foreigners tweets about the post conflict in Colombia, which began at the end of the year 2016 with the covenant established between the Colombian Government and the FARC-EP, the main subversive group of this country. The post conflict can be defined as the period of time that began with the culmination of an armed conflict.

This paper concludes that Foreigners have a better perception of the post conflict that the Colombians. The results was obtained a percentage of 60% expressing positive feelings of Foreigners in contrast with the Colombians who only had a 20%. The good perception of Foreigners on Colombia in security aspects can help increase tourism substantially, which helps to improve the economy of the country and social development.

2. COLOMBIAN POSTCONFLICT

The FARC-EP's army was a revolutionary movement of political-military character born in the year of 1964 in the south of the Tolima department. The armed conflict in Colombia has been one of the longest in the world and one of the most fatalities has left in its wake [2]. The conflict between FARC-EP rebels vs. Colombian government is one of the oldest in western countries, including thousands of deaths, massacres and terrible living conditions [3]. The conflict in Colombia has been developed in three stages: initial, intermediate and final. The final stage known as the post conflict, is the longest, costly and that has a greater number of challenges [2].

In November of 2016, after a confrontation more than fifty years, the Colombian Government and FARC-EP agreed to put a definitive end to the internal armed conflict. This new social condition brought many benefits for the country, since agriculture has been renewed and trade and tourism have been activated in many regions that were affected by this

senseless war [3].

The Colombians have been witness to on-going peace talks between the government and the FARC-EP, guerrilla group, a whole set of practices of forgiveness, inaugurations of memory museums, the passing of the victims and land restitution law, as well as the recognition of victims and the effervescence of their movements and organizations. These factors have spiralled victims' demonstrations and fuelled marches all over the country [4].

There are increased rates of foreign investment in those regions that, until as little as a decade ago, were subject to appalling rates of internal displacement, massacres and disappearances, and which now constitute new borders where global capitalism has been able to reconfigure and increase its rates of capital accumulation [4]. In the post conflict context, effective peace education has a stronger practical than didactic focus, and it stimulates empathy, cooperation, reconciliation, and community processes for handling conflict in a nonviolent manner. These processes play a key role also in the prevention of children's engagement in violence and terrorism [4].

3. BACKGROUND AND RELATED WORK

Sentiment analysis is one of the fastest growing research areas in sciences, making it challenging to keep track of all the activities in the area [5]. This seeks to identify the viewpoints underlying a text span [6]. This aims to automatically uncover the underlying attitude that we hold towards an entity. The aggregation of these sentiment over a population represents opinion polling and has numerous applications [7]. Sentiment analysis and opinion mining is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. It is one of the most active research areas in natural language processing and is also widely studied in data mining, web mining and text mining [8]. Sentiment is a ubiquitous and constant part of the human life, with variations in sentiments changing only slightly up or down [9], the sentiments are episodes of coordinated changes in response to events of major significance [10].

In the literature review we found researches that apply sentiment analysis, in [5] investigated the public opinions and sentiments towards the Syrian refugee crisis, which has affected millions of people and has become a widely discussed, polarizing topic in social media around the world. In [11] developed a sentiment analysis of the Twitter discussion on the 2016 Austrian presidential elections, they extracted and analyzed a data-set of Twitter messages. In [12] developed a multifacet sentiment analysis approach to measure the dimensions in consumer reviews, they were interested in the comments on product quality and service quality. Figure 1, presents the stages for the feelings classification [13]:
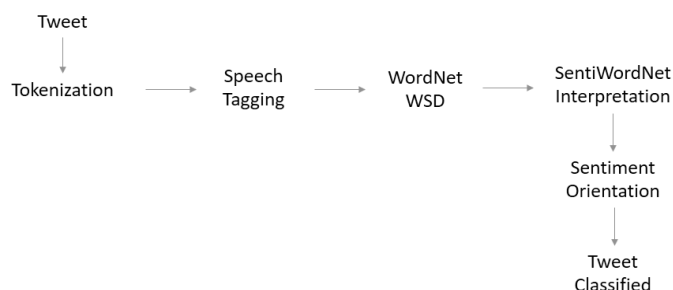

Figure 1: Classification feelings phases. Source: adapted from [13]

- *Tokenization*: splits the text into very simple tokens such as numbers, punctuation and words of different types.
- *Speech tagging*: produces a tag as an annotation base on the role of each word in the tweet
- *WordNet WSD*: determinate the meaning of every word in his context
- *SentiWordNet interpretation*: given a synset we can search the sentiment score associated to the synset.
- *Sentiment orientation*: term score summation method. The positive and negative scores for each term found in a tweet are summed separately.

Text mining is the process of analyzing collections of text materials in order to capture the themes and key concepts and discover the hidden relationships and trends without the need to know the exact words or terms that the authors have used to express these concepts. Text mining and the action of retrieving information are concepts that are sometimes confused, although they are quite different [14].

4. MATERIAL AND METHODS

4.1 Data Collection

In the analysis, 500 tweets of Colombian and Foreigners of the social network "Twitter" were used. This is one of several social media networks, and there is little reason to suspect that data from Twitter is fundamentally different from

other socially generated data [9]. The search of tweets was conducted been 01/06/2017 to 01/09/2017, the tweets was saved in .xlsx files in preparation for the analysis process. Colombian and Foreigners tweets were analyzed separately. For Colombian Tweets 250 tweets were gathered with the keywords: "FARC-EP", "Colombia" and "Post conflicto". For Foreigners tweets, we used the keywords: "Post conflict", "Colombian" and "Peace" for the search. In total, 250 Foreigners tweets were gathered. For the analysis and tweets classification was used the Excel tool "TEXT2DATA", a simple and powerful tool for Analysts and BI developers [15]. This tool offers text analytics SaaS services like in-depth analysis, social media data, cloud computing and features such as sentiment analysis API, Twitter sentiment engine, user classification models [15]. This tool is downloaded from the following link: http://text2data.org/Excel, is installed in MS Excel and runs directly from the Developer tab, as presented in Figure 2:
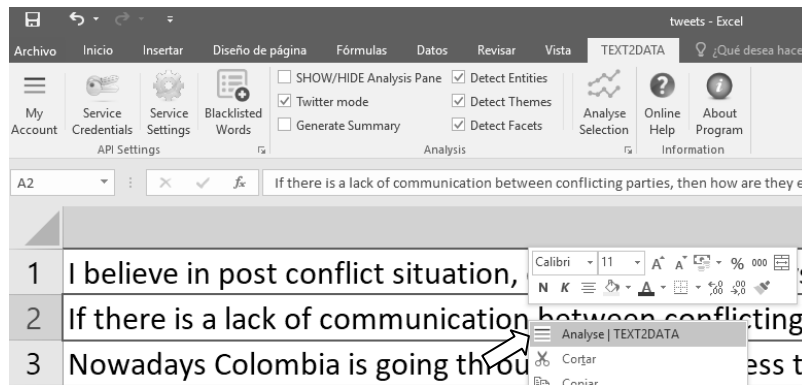

Figure 2: Text2Data Tool. Source: author elaboration

For the analysis were deleted the comments that had double meaning, abbreviations and illogical ideas. The tweets duplicates were eliminated. Table 1, presents some examples of tweets used in the analysis:

Table 1: Post conflict tweets

User	*Comment*
@christerMyklebu	"I believe in post conflict situation; communication plays more than a mere logistic role. Communication in this cases may serve to infuse hope and breathe life to a community. It is a tool for bonding and being aware of that there's light at the end of the tunnel"
@a_martinezgomez	"If there is a lack of communication between conflicting parties, then how are they ever going to get out? Here they need to learn to live with each other again - communicate, share, interact, move on"
@altamar_co	"Nowadays Colombia is going through a peace process that is why the post conflict is the perfect opportunity to get to know this beautiful country that has virgin beaches, incredible sunsets and unique landscapes to the nature photography"

4.2 Data Pre processing

In the data cleaning, the retweets were not considered, the aim of this research is to specify the opinions of people (Colombians and Foreigners). The uppercase letters were changed into lowercase characters. Punctuation, hyperlinks and numbers were deleted from the tweets and only words were considered for the analyses. The tweets were split into tokens and those words were looked up in each tweet and scores were calculated. The scores of the tweets were aggregated based on the token sentiment scores.

4.3 Calculation of term frequencies

Word cloud is a useful tool for visually a large amount of data [16]. In this paper, word clouds were generated for visualize the text data of the Colombian and Foreigners tweets. These, were created by using "Text2Data", a MS Excel tool for create word clouds. Word clouds is a great way of visualizing a piece of text or a news feed. The cloud gives greater prominence to words that appear more frequently in the tweets [17].

5. RESULTS AND DISCUSSION

"TEXT2DATA", classifies tweets into five categories: very negative, negative, neutral, positive and very positive. The Colombian tweet dataset contained 250 tweets. For each tweet an overall sentiment score was generated. Then, the tweets were grouped into the sentiment categories. Out of 250 Colombian tweets, 130 of them were classified as negative, 25 as neutral, 50 as positive, 30 as very negative and only 15 tweets were classified as very positive. Figure 3, presents the sentiment analysis results of Colombian tweets. Negative neutral tweets dominated the sentiment categories.

After, positive tweets were the second in order and followed by the very negative tweets. Figure 4, presents the pie chart of the analyzed Colombian tweets. The proportions of the results composed negative with the highest rate of 52%, 20% as positive, followed by 12 % as very negative, respectively. 10 % of the tweets were categorized as neutral and only 6 % as very positive. The results show that Colombian speaking community in twitter post tweets mostly negative opinions regarding the topics of post conflict.
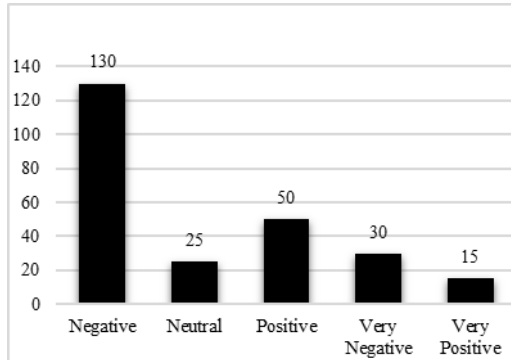

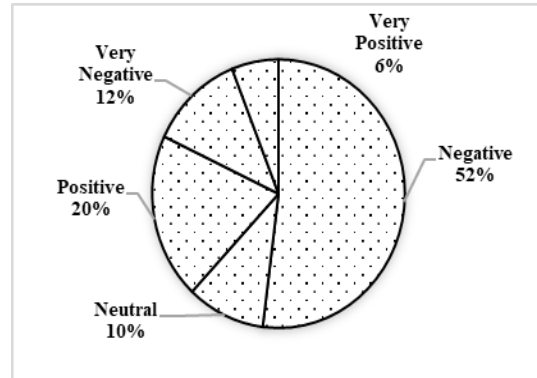**Figure 3.** Colombian tweets result graph

Figure 4. Colombian tweets results pie chart

Out of 250 foreigner's tweets, 150 of them were classified as positive, 40 as very positive, 35 as negative, 15 as neutral and only 10 tweets were classified as very negative. Figure 5, presents the sentiment analysis results of Foreigners tweets. Positive tweets dominated the sentiment categories. After, very positive tweets were the second in order and followed by the negative tweets. Figure 6, presents the pie chart of the analyzed Colombian tweets. The proportions of the results composed positive with the highest rate of 60%, 16% as very positive, followed by 14 % as negative, respectively. 6 % of the tweets were categorized as neutral and only 4 % as very negative. The results show that Foreigners speaking community in twitter post tweets mostly positive opinions regarding the topics of Colombian post conflict.
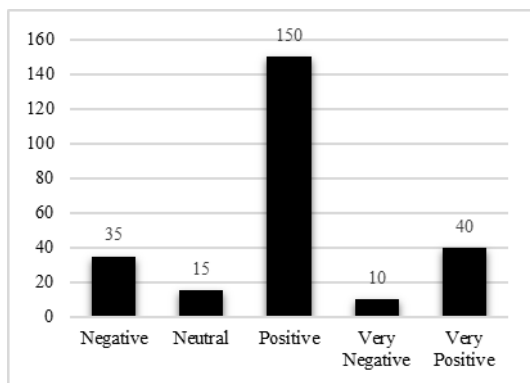

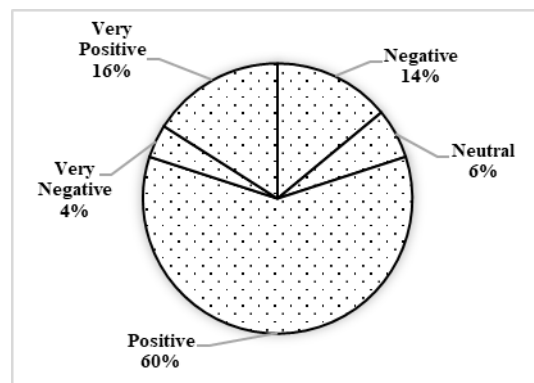**Figure 5.** Foreigner's tweets analysis graph

Figure 6. Foreigner's tweets analysis pie chart

5.1 Word clouds of frequent terms

This work analyzed the word frequencies of Colombian and Foreigners tweets about post conflict using word clouds. Figure 7, presents the word cloud results for Colombian tweets.

Figure 7. Word Cloud for Colombian tweets

The word "conflict" was the most common word among the terms in Colombian tweet data. This term is the boldest word in the cloud, it is followed by the word "Colombia". Figure 8, presents the word cloud results for Foreigners tweets. The most common word was "Environmental", it is followed by the word "Conflict".

Figure 8. Word cloud for foreigner's tweets

As observed on the word cloud graphic of the frequent terms, the Colombian tweets often involved discussions regarding the government, massacres, deaths and illegal armed forces, the Colombian tweets focused more on the information about the violations of the peace agreement established with the FARC-EP. In Foreigners tweets the words "peace", "resources", "economic", "communities", "biodiversity", "agricultural" were counted as frequent. In contrast to Colombian Tweets, the Foreigners tweets seemed to focus more on the environment and natural resources. There were some words frequently used in both datasets. These words were mostly related to the conflict. For instance, "drugs",

"victims", "armed", "impact" and "damages". The list of terms that are used in the analysis is shown by category and language in Table 2 and Table 3:

Table 2. Colombian terms

Humanity	War	Politics
Victims	Deaths	Government
Population	Homicide	Politics
Civilians	Killing	Agreement
Children	Displacement	Laws
Men	Damages	State
Women	Armed	Congress
Ancients	Unemployment	Illegal
Work	Assassin	Post conflict
President	Impact	Money
Communities	Drugs	
	Narcotraffic	
	Corruption	
	Damages	

Table 3. Foreigners terms

Humanity	War	Politics
People	Dead	Legal
Resources	Steal	Agreement
Tourism	Damages	Restitution
Agricultural	Marijuana	Country
Child	Cocaine	City
Resources	Illicit	Town
Tourist	Damages	Department
Biodiversity	Cultivation	Norm
Income	Armed	Economic
Agricultural	Impact	
	Extortion	
	Drugs	
	Peace	

6. CONCLUSIONS

The analysis of results allowed identifying that Foreigners have a better perception of the post conflict that Colombians, a situation that can benefit the tourism and economic growth, by the increase in foreign investment. On the other hand, the bad perception of the Colombians can be justified by the discontented and nonconformities with some established agreements with the FARC-EP, among which are the adoption of the Victims Law, land restitution and some practices of forgiveness that left in impunity a large number of innocent people deaths.

Sentiment analysis allows to determine the satisfaction or dissatisfaction degree of comments made by people, facilitating the identification of aspects that must be improved, which can help to take measures that would contribute to the solution of social problems such as violence, poverty and social inequality. Specifically, in this work, helped to identify that foreign investment can be increased in the country, due to the good perception of Foreigners on post conflict in Colombia.

7. REFERENCES

[1] I. P. Windasari, F. N. Uzzi, and K. I. Satoto, "Sentiment Analysis on Twitter Posts: An analysis of Positive or Negative Opinion on GoJek," pp. 266–269, 2017.

[2] J. C. Rojas, "Etapas del conflicto armado en Colombia: hacia el post conflicto," Latinoamérica. Rev. Estud. Latinoam., vol. 62, pp. 227–257, 2016.

[3] A. Higuera and J. D. Ramírez, "The Colombian peace deal and its impact on the evolution of tropical diseases agents," Infect. Genet. Evol., vol. 57, no. October 2017, pp. 145–150, 2018.

[4] H. Kupermintz and G. Salomon, "Lessons to Be Learned from Research on Peace Education," Theory Pract., vol. 44, no. 776129590, pp. 293–302, 2005.

[5] N. Öztürk and S. Ayvaz, "Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis," Telemat. Informatics, vol. 35, no. 1, pp. 136–147, 2017.

[6] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," in Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics, 2004.

[7] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S. F. Chang, and M. Pantic, "A survey of multimodal sentiment analysis," Image Vis. Comput., vol. 65, pp. 3–14, 2017.

[8] B. Liu, "Sentiment Analysis and Opinion Mining," Synth. Lect. Hum. Lang. Technol., vol. 5, no. 1, pp. 1–167, 2012.

[9] N. Danneman and R. Heimann, Social Media Mining with R. Birmingham: Packt Publishing Ltd, 2014.

[10] K. R. Scherer, "Psychological models of emotion.," in The neuropsychology of emotion., New York, NY, US: Oxford University Press, 2000, pp. 137–162.

[11] E. Kušen and M. Strembeck, "An Analysis of the Twitter Discussion on the 2016 Austrian Presidential Elections," Online Soc. Networks Media, vol. 5, pp. 37–50, 2017.

[12] T. P. Liang, X. Li, C. T. Yang, and M. Wang, "What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach," Int. J. Electron. Commer., vol. 20, no. 2, pp. 236–260, 2015.

[13] P. Perner, Advances in Data Mining: Applications and Theoretical Aspects, 9th ed. Leipzig: SpringerLink, 2009.

[14] IBM Knowledge Center, "Acerca de la minería de textos," 2018. [Online]. Available: https://www.ibm.com/support/knowledgecenter/es/SS3RA7_16.0.0/com.ibm.spss.ta.help/textmining/shared_entities/ tm_intro_tm_defined.htm. [Accessed: 16-Jan-2018].

[15] Text2Data, "Introducing Sentiment Analysis and Text Analytics Add-In for Excel," Tool for Analysts and BI developers, 2018. [Online]. Available: http://text2data.org/Excel. [Accessed: 14-Feb-2018].

[16] N. Öztürk and S. Ayvaz, "Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis," Telemat. Informatics, vol. 35, no. 1, pp. 136–147, 2018.

[17] Microsoft, "Pro Word Cloud," 2018. [Online]. Available: https://appsource.microsoft.com/en-us/product/office/wa104038830?src=office&corrid=1ffc7108-1353-4d77-b7e4-9d56cad7493f&omexanonuid=91aad751-c003-4a30-bf58-a392a3d2ab0d. [Accessed: 14-Feb-2018].