THE FEMICIDE IN COLOMBIA AND MEXICO: A TEXT MINING ANALYSIS

V. D. Gil[1], J. D. Betancur[1], I. C. Puerta[1], L.M. Montoya[2], J.M. Sepulveda[2]

[1] NBA "Universidad Católica Luis Amigó", Transversal 51A # 67B 90 Medellín, Colombia

[2] INGENIAR "Corporación Universitaria UNIREMINGTON", Calle 51 No. 51-27 Medellín, Colombia

ABSTRACT
In the last decade the femicide has become one of the most frequent types of killings in the world, especially in Latin American countries. Despite the great strides to achieve gender equity, violence against women persists in society. The objective of this paper was to perform a tweets analysis of feelings and opinions from the social network Twitter related to the femicide in Colombia and Mexico, which has become a widely debated topic in the world. 300 tweets of Colombians and 300 tweets of Mexicans were collected. The opinions of Mexicans 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 different from Mexicans. The tweets of Mexicans have more negative feelings compared to the Colombians tweets. In comparison with the proportion of tweets with negative feelings of Mexicans, 83%, the percentage of Colombians tweets is 74%.

Keywords: femicide, text mining, Mexico

INTRODUCTION

At present the femicide has become one of the most frequent types of killings in the world, especially in Latin American countries. Despite the great strides to achieve gender equity, violence against women persists in society. The femicide is one of the most recurrent problems of the XXI century. Thousands of women are exposed to all kinds of violence to which, in many cases, they end up murdered by their sentimental couples or strangers. The most worrying aspect of this situation is that the murder of a woman became something natural and the femicide does not appear to be penalized with any kind of moral or social guilt.

The objective of this paper was to perform a sentiment analysis of Colombians and Mexicans tweets about the femicide in the world. For the analysis, 600 tweets of Colombians and Mexicans of the social network "Twitter" were used. The tweets were analyzed separately. For the analysis and tweets classification the statistical software "R 3.4.3" was used, a powerful tool for data analysis.

This paper concluded that Colombians and Mexicans have a bad perception of femicide. The results were obtained a percentage of 74% expressing negative feelings of Colombians and the Mexicans results was 83%. The bad perception about this problematic is due to the lack of laws and penalties for this crime, the femicide has been found in the impunity to your best ally. The governments of the Latin America countries should establish laws that severely punish this crime, which takes more force in the world.

Femicide
The femicide or murder of women is a public health problem and a serious violation of human rights [1]. In Latin America, homicide is a leading cause of death among women [2]. Currently it is not known with accuracy the magnitude of this problem in the world, due to the methodological difficulties to differentiate these killings of other types of women homicides. The term femicide, used by Diana Russell in 1970, makes reference to the murders of women for gender reasons [1]. This term is to differentiate the crimes of gender violence of the women murders for other reasons. The term includes the murders committed by the partner or ex-partner, honor crimes (murder of women committed by family members

to restore the family honor), homicides associated with rape and prostitution, and the selective homicides of women during the wars [2]. Homicide is the consummation of a long process in which the women is subjected to some type of violence, which empties into the femicide. The author of the crime can be a man or woman, however, men's violence against women has been shown to be more massive, severe, and capable of generating large social impacts [3]. Femicide has four causes: badly managed male aggressivity; slavery in human relationships; various forms of jealousy and emotional dependency [4]. Other possible social causes of femicide are: rapid population growth and changing demographic structure, poverty and unemployment [5]. The acts of violence by their sentimental couple in Colombia reported between January and October of the year 2017 reach 35.690 cases [2]. These quantities reveal the magnitude of the problem, since a large percentage of these women could be killed if measures are not taken appropriate legal and socio-cultural strategies. The femicide affects the family nucleus, because if the victim has sons and the murder is carried out by their sentimental partner, they could lose their two parents and be socially stigmatized being the sons of the murderer of his mother. The possibility of femicide increases with the presence of sons of past sentimental relationships [6], this is only one of the possible scenarios in which there may be a femicide.

From a cultural perspective, the murderer in the majority of cases kills women for economic or sentimental reasons [7]. Many research has indicated that the most frequent perpetrators of intimate femicide are husbands, these findings indicate that persons who are separated are the most frequent perpetrators of intimate femicide [8]. The power factor can lead a man to become a femicide, as well as some Japanese soldiers during the second world war, raped girls and women; many of these women died due to sexually transmitted diseases, or because they didn't want to provide the service imposed by the imperial forces [9]. From a biological perspective little research has been done on the factors that characterize a femicide. These have been studied as any other type of killers, which makes it difficult to analyze the femicide as a killer different than other types of killers. From the neuropsychology causes have been studied to the femicide to identify brain areas that suffer alterations [10]. One of the ways in which you could understand the biological factors that motivate a femicide, would be to assess the true impact of risk factors like suffering a mental illness, alcohol, hallucinogens or carry weapons. Filing a judicial complaint against the aggressor does not seem to affect the risk of murder among women exposed to domestic violence. Protection measures for women should be strengthened among immigrant women and women living in urban and rural areas [11]. The elements of the legal culture of society performed the task of creating the legal system such as become the sphere of social experimentation [12].

BACKGROUND AND RELATED WORK

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. Sentiment analysis is a branch of the text mining; this consists in the application of a range of technologies to determine sentiments expressed within social media about particular topics [13]. This is one of the fastest growing research areas in sciences, making it challenging to keep track of all the activities in the area [14]. This seeks to identify the viewpoints underlying a text span [15].

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 [16]. 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 [17]. Sentiment is a ubiquitous and constant part of the human life, with variations in sentiments changing only slightly up or down [18], the sentiments are episodes of coordinated changes in response to events of major significance [19].

In the literature review we found research that apply sentiment analysis; Gan et al. [20] developed a research for to identify the structure of online restaurant reviews and examine the influence of review attributes and sentiments on restaurant star ratings. Other authors determined whether the results obtained by applying sentiment analysis to data extracted from social media can reveal the political preferences of

citizens to a greater degree of accuracy than traditional public opinion surveys [21]. In Öztürk and Ayvaz [14] 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. Kušen and Strembeck [22] developed a sentiment analysis of the Twitter discussion on the 2016 Austrian presidential elections, they extracted and analyzed a data-set of Twitter messages. Liang et al. [23] developed a multifaceted sentiment analysis approach to measure the dimensions in consumer reviews, they were interested in the comments on product quality and service quality. Kavitha and Usha [24] used to perform intrusion detection in Wireless Local Area Networks to achieve more accuracy in intrusion detection. Thanigaivel and Kumar [25] identified hidden patterns by applying data mining techniques, which are noteworthy to heart diseases. The stages for the feelings classification are [26]: 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.

MATERIAL ANS METHODS
Data collection
Actually, the information and communication technologies (ICT) showed great potential for enhancing and merging social networks [27]. Social networks are growing day by day. The users of the these are generating values for these networks [28]. Actually, the Internet usage rates have increased in the world number of internet users reached to 3 billion and half of them have active user status on social media [29]. In the analysis, 600 tweets of Colombian and Mexican of the social network "Twitter" were used. At present, this is one of several social media networks. The search of tweets was conducted between 11/01/2018 to 11/02/2018, the tweets was saved in csv files in preparation for the analysis process. The tweets of the Colombian and Mexican were analyzed separately. For Colombian Tweets, 300 tweets were gathered with the keywords: "Femicide", "Colombia" and "Women". For Mexicans tweets, we used the keywords: "Femicide", "Mexico" and "Women" for the search. In total, 300 Mexican tweets were gathered.

The tweets comply with the four characteristics of the data analysis, known as the 4Vs: volume, velocity, variety and veracity [30]. For the analysis and tweets classification was used the statistical software "Rcran 3.4.3", a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS [4]. Was used the packages: "tm", "SnowballC", "wordcloud" and "RColorBrewer". For the transformation and cleaning process was used the "m_map()" function and the arguments: "removeNumbers", "removePunctuation" and "stripWhitespace". Fig. 1:

```
# Install
install.packages("tm")  # for text mining
install.packages("SnowballC") # for text stemming
install.packages("wordcloud") # word-cloud generator
install.packages("RColorBrewer") # color palettes
# Load
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
```

Fig. 1: Packages and libraries used for the analysis

Data preprocessing
For the analysis the comments were deleted that had double meaning, abbreviations and illogical ideas. The tweets duplicates were eliminated. In the data cleaning, the retweets were not considered, the aim of this research was to specify the opinions of people (Colombians and Mexicans). 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. Table 1, presents some examples of tweets used in the analysis:

Table 1
Feminicide tweets

User	Comment
@hansiipinge	Yes It is definitely skewed towards women So much so that we coined terms liken femicide Just because we freaking could screw men
@ Bouncer Mangel	What they did was femicide Now they have produced robots to provide companionship Robots with white girly looks which also help in home shores
@KhSophia	My life these days is basically discussing femicide in the mornings and doing fashion shows in the evenings

Calculation of term frecuencies

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 Mexicans tweets. 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].

RESULTS AND DISCUSSION

The tweets were classified into five categories: very negative, negative, neutral, positive and very positive. The Colombian tweet dataset contained 300 tweets. For each tweet an overall sentiment score was generated. Then, the tweets were grouped into the sentiment categories. Out of 300 Colombian tweets, 222 of them were classified as negative, 15 as neutral, 7 as positive, 54 as very negative and only 2 tweets were classified as very positive. Fig. 2, presents the sentiment analysis results of Colombian tweets. Negative tweets dominated the sentiment categories. After, very negative tweets were the second in order and followed by the neutral tweets. Fig. 3, presents the pie chart of the analyzed Colombian tweets. The proportions of the results composed negative with the highest rate of 74% followed by 18% as very negative, 2% as positive respectively. 5% of the tweets were categorized as neutral and only 1% as very positive. The results show that Colombian speaking community in twitter post tweets mostly negative opinions regarding the topics of femicide.
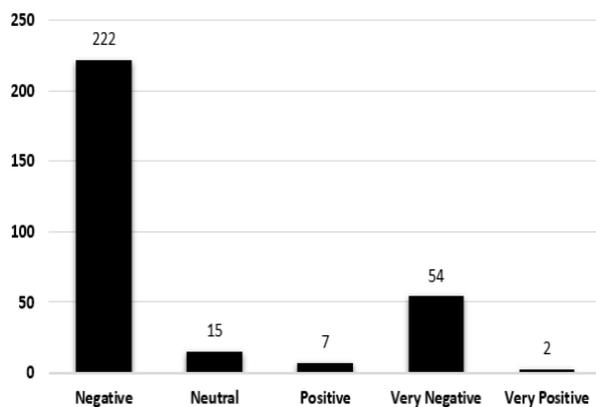

Fig. 2: Colombian tweets results graph
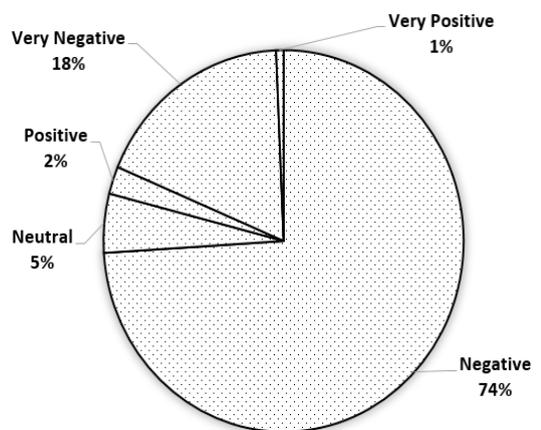
Fig. 3: Colombian tweets results pie chart

Out of 300 Mexican tweets, 10 of them were classified as positive, 3 as very positive, 250 as negative, 7

as neutral and 30 tweets were classified as very negative. Fig. 4, presents the sentiment analysis results of Mexican tweets. Negative tweets dominated the sentiment categories. After, very negative tweets were the second in order and followed by the positive and neutral tweets. Fig. 5, presents the pie chart of the analyzed Mexican tweets. The proportions of the results were composed negative with the highest rate of 83%, 1% as very positive, 3% as positive and 3% as neutral respectively and 10% as very negative. The results show that Mexican speaking community in twitter post tweets mostly negative opinions regarding the topics of femicide.

Fig. 4: Mexican tweets results graph

Fig. 5: Mexican tweets results pie chart

Word clouds of frequent terms

This paper analyzed the word frequencies of Colombian and Mexican tweets about femicide using word clouds. Fig. 6, presents the word cloud results for Colombian tweets.

Fig. 6: Word Cloud for Colombian tweets

The word "women" 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 "femicide" and "violence". Fig. 7, presents the word cloud results for Mexican tweets. The most common word was "women", it is followed by the word "Mexico".

As observed on the word cloud graphic of the frequent terms, the Colombian tweets often involved discussions regarding the deaths and social problems, the Colombian tweets focused more on the information about the murder of women in the country. In Mexican tweets the words "women", "murdered", "violence", "Mexico" were counted as frequent. There were some words frequently used in both datasets. These words were mostly related to the femicide; for instance, "women", "violence" and "femicide". The list of terms that are used in the analysis is shown by category and language in Tables 2 and 3:

Fig. 7: Word cloud for Mexican tweets

Table 2
Colombian terms

Geographic	Violence	Social
Colombia	Homicide	Impunity
America	Femicide	Inequality
International	Murder	Rights
Latin	Death	Institutional
Cali	Violent	Judicial
National	Oppression	Society

Table 3
Mexican terms

Geographic	Violence	Social
Mexico	Dead	Justice
Morelos	Killed	State
Capital	Raped	Cartels
America	Femicide	Federal
Chihuahua	Crime	Government
National	Murders	Law
Municipali	Death	Authors
		Women

Research Article - This article was checked by Turnitin

175

	Torture Cruelty Abuse	Public Security Human	ty Ecatepec Country	Perpetrated De manded	

CONCLUSION

The analysis of results allowed identifying that Colombian and Mexican people have a bad perception of the femicide, a situation that can help to raise awareness and to decrease the number of innocent women murders. The bad perception of these is due to the terrible consequences that are generated by the society for the death of mothers with sons, a situation that destroys the homes and affects the normal development of boys and girls in the world. 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 femicide, violence and social problems. Specifically, in this work, helped to identify that from the family nucleus is failing to shape rational men and women, it is not right that women are killed by those who said love ever.

ACKNOWLEDGEMENT

REFERENCES
Sanz-Barbero, B. Otero-García, L. Boira, S. Marcuello, C. Cases, C. V. Acción COST Femicide Across Europe, un espacio de cooperación trasnacional para el estudio y el abordaje del feminicidio en Europa. Gac. Sanit, 2016. Volume 30. Issue 5. pp. 393–396.

Murillo, F. H. S. Olmo, J. C. de Cortázar, A. R. G. The spatial heterogeneity of factors of feminicide: The case of Antioquia-Colombia, Appl. Geogr, 2018. Volume 92. pp. 63–73.

Acero, A. Aproximaciones a los conceptos de femicidio, feminicidio y homicidio en mujeres", 2009. Web. 23 Feb, 2018 <http://www.medicinalegal.gov.co/documents/20143/49505/Homicidio.pdf>

Pasini, W. Le féminicide : aspects psychodynamiques, Sexologies, 2016. Volume 25. Issuee. pp. 7–10.

Ikechukwu, G. Science, P. Urbanization and Insecurity in Nigeria: The Issues, Challenges and Prospect for National Development. Middle-East J. Sci. Res, 2018. Volume 26. Issue 1. pp. 68–77.

PAHO. Femicide Understanding and addressing violence against women, Feminicide, 2012. Web. 23 Feb, 2018. <http://apps.who.int/iris/bitstream/10665/77421/1/WHO_RHR_12.38_eng.pdf?ua=1>

Laurent, C. Femicide The Killing of Women and Girls Around the World, Academic Council on the United Nations System (ACUNS) Vienna Liaison Office, 2013. Web. 23 Feb, 2018. <https://acuns.org/wp-content/uploads/2013/05/Claire-Laurent.pdf>

Stout, K. Intimate Femicide, J. Offender Rehabil,1993. Volume19. Issue 3. pp. 81–94.

Lamas, M. Feminismo y prostitución: la persistencia de una amarga disputa, Debate Fem, 2016. Volume 5. pp. 18–35.

Pedraza, S. Escobar, Lady. Gonzalez, S. Evaluación Neuropsicológica Compleja del Delito de Homicidio, Umbral científico, 2008. Volume 13. Issue 1. pp. 162.

Sanz-Barbero, B. Heras-Mosterio, J. Otero-García, L. Vives-Cases, C. Perfil sociodemográfico del feminicidio en España y su relación con las denuncias por violencia de pareja, Gac. Sanit, 2016. Volume 30. Issue 4. pp. 272–278.

Kovalenko, K. E. Features of the Development of Justice, Middle-East J. Sci. Res, 2014. Volume 22. Issue 2. pp.176–179.

Kennedy, H. Perspectives on Sentiment Analysis, J. Broadcast. Electron. Media, 2012. Volume 56. Issue 4. pp. 435–450.

Öztürk, N. Ayvaz, S. Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis,

Telemat. Informatics, 2018. Volume 35. Issue1. pp. 136–147.

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

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

Liu, B. Sentiment Analysis and Opinion Mining, Synth. Lect. Hum. Lang. Technol, 2012. Volume 5. Issue 1. pp. 1–167.

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

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

Gan, Q. B. Ferns, H. Yu, Y. Jin, L. A Text Mining and Multidimensional Sentiment Analysis of Online Restaurant Reviews, J. Qual. Assur. Hosp. Tour, 2017. Volume 18. Issue 4. pp. 465–492.

Oliveira, D. J. S. de Souza Bermejo, P. H. dos Santos, P. A. Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls, J. Inf. Technol. Polit, 2017. Volume 14. Issue 1. pp. 34–45.

Kušen, E. Strembeck, M. An Analysis of the Twitter Discussion on the 2016 Austrian Presidential Elections, Online Soc. Networks Media, 2017. Volume 5. pp. 37–50.

Liang, T. P. Li, X. Yang, C. T. Wang, M. What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach, Int. J. Electron. Commer, 2015. Volume 20. Issue 2. pp. 236–260.

Kavitha, P. Usha, M. Time Orient Flow Estimation Based Data Mining Approach for Intrusion Detection in Wireless Local Area Networks Using Delay Averaging Scheme Department of Computer Science and Engineering, Middle-East J. Sci. Res, 2016. Volume 24. Issue 1. pp. 97–102.

Thanigaivel, R. Kumar, R. Boosted Apriori: an Effective Data Mining Association Rules for Heart Disease Prediction System, Middle-East J. Sci. Res, 2016. Volume 24. Issue1. pp. 192–200.

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

Hameed, T. M. Bin, Z. Hassan, H. Sulaiman, R. Is Social Network an Effective E-learning Tool: A Survey, Middle-East J. Sci. Res, 2015. Volume 23. Issue 1. pp. 119–126.

Sindhu, R. Gehlot, N. Analysis and Quantitative Approach for Measuring the Value of the Social Network, Middle-East J. Sci. Res, 2018. Volume 26. Issue 1. pp. 60–6

Keçe, M. Facebook Addiction Levels of Secondary Students: A Scale Development Study, Middle-East J. Sci. Res, 2016. Volume 24. Issue 2. pp. 214–226.

Gil, V. Learning Analytics and Scholar Dropout: A Predictive Model, Middle-East J. Sci. Res, 2017. Volume 25. Issue 7. pp. 1414–1419.