

Public Safety Perception in Ecuador: An Approach from Social Networks over Data Analytics

Maria de Lourdes Díaz, Jorge Berrezueta, Gonzalo Albán Molestina and Andres Ortega*

Universidad Ecotec, Samborondón, Ecuador

Abstract

In Ecuador, insecurity and crime address a space that generates a lot of commotion in social networks. The data provided by the governments of the nations is not contrasted with what happens in public opinion. Today, information is very sensitive thanks to the use of social networks, where it is sought through a data analytics tool to measure the perception of insecurity of citizens. Based on the metadata offered by the Twitter API, we collect this information through an algorithm based on natural language processing (NLP) using Python, we generate a statistical report to understand the context of citizen perception. The results show that there is a high correlation of security factors such as theft, corruption, crime at the regional and territorial level, affecting the cultural, political and economic development of cities and countries.

Keywords

Social Networks, Data Analytics, Homicides, Public policies, NLP

1. Introduction

Social media allows users to create a profile, navigate, connect, and communicate with other users through private or public messaging [1]. At the same time, social media provides a space that was originally designed to be a thought gatherer. This kind of platform was made to be entertaining, having algorithms fine-tuned to display relevant content to each user based on their previous history and interactions with content by other users. Over time, social media experimented a growing shift toward other purposes of use, where they gained trust and market share as the primary news source for many users where content is usually not filtered and does not exclude subjective thoughts from other users on a certain topic or situation [2].

Information on the social network Twitter is abundant and available, as well as giving rise to fresh opinions on current contexts [3]. Keep in mind that on this platform users not only post tweets, but can also receive responses or interactions. What is known as "retweeting" is a way of spreading a message regardless of its veracity that users can use as a method of interaction. Frequently, this interaction is found in messages of social and political information,

ICAIIW 2022: Workshops at the 5th International Conference on Applied Informatics 2022, October 27–29, 2022, Arequipa, Peru

*Corresponding author

✉ mariadiaz@est.ecotec.edu.ec (M. d. L. Díaz); joberrezueta@est.ecotec.edu.ec (J. Berrezueta); galban@ecotec.edu.ec (G. Albán Molestina); aortegao@ecotec.edu.ec (A. Ortega)

ORCID 0000-0002-9141-2048 (A. Ortega)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

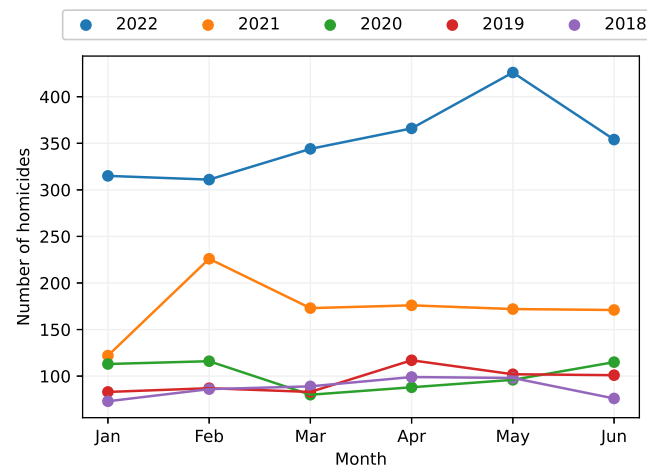


Figure 1: Semi-annual homicides in Ecuador

news, opinions and very commonly controversial topics to generate a representation of the current perception of people [4].

This research starts by categorizing crime, violence and delinquency as synonymous words that affect citizen security, words that indicate a deviation in the behavior of individuals within a society as well as the violation of established rules and codes [5]. Latin America is considered the most violent region in the world and its origin may be relevant to factors such as the rapid conversion from rural to urban, producing a standard of living that requires large investments, inefficient public social services and evident economic inequality [6].

Insecurity in Ecuador is a social problem that afflicts all its citizens. There are different forms of citizen insecurity that surround the country. Criminal events are internalized by not obeying legal and moral norms. Exposure to the use of violence generates an expansion of criminal acts as a way of solving the problems that society may be going through.

In the last 5 years, they have focused mainly on the number of robberies and homicides. According to data reported by the Ministry of Government on its information portal the six months statistic can be perceived in the last 5 years. The accuracy presented in Figure 1 allows visualizing this evolution of incidents which show that between 2021 and 2022 there have been a greater number of intentional homicides. However, the reported data is usually not updated. For example, the State Attorney General's Office which receives complaints about events of this type, filed a last report of homicides and robberies for the period of "January-November 2021".

The impact of the insecurity crisis is due to the deterioration of the quality of life, this type of difficulties gives a new perception of the current state of security by the victims who face a transformation of habits and in their daily routines in order to avoid being involved in a dangerous situation [7].

Insecurity is associated with fear and concern, mainly affecting the calm for a citizen to function smoothly within a society. Fear of crime and feelings of insecurity regarding one's environment have an impact on the quality of life of citizens, especially when feelings of fear become excessive with consequences that can lead to health problems, as it is argued that

general anxiety is significantly related to fear of crime [8]. These factors contribute to the fact that citizens limit their decision-making in multiple aspects such as consumption, investments and mobility, directly affecting the social and economic development of our country. [9].

The perception of security is a variable that subjectively indicates the concern among citizens since it is seeking to measure the fear of danger that Ecuadorians feel day to day. But beyond perceptions, there is no clarity on the existence of tools or strategies to follow to reduce and control insecurity. This data can give citizens a certain notion of security about what is happening in the country [10].

However, this does not invalidate the evident increase in the number of crimes, which generates a progression of the perception of insecurity with respect to victimization. Ecuador carried out a single survey on the perception of insecurity at the national level in 2011. This data by not being updated causes numerous unknowns about the current levels of victimization and generates distrust in public institutions in charge of security, such as the Ecuadorian police [11]. There is a study in Mexico that takes bases analyzing the relationships between victimization, perception of insecurity and changes in routines through an adaptation of the National Survey of Victimization and Public Security, this pointed out important data for the development of the subject, for example, that women and men victims of crime indicated restrictions in daily life [7].

How can we measure the perception of security in the country with available resources? Currently there is a lack of a national methodology for measuring the perception of insecurity through the use of information on social networks. The content that travels on social networks affects directly or indirectly the perception that people have about a wide variety of topics. Although this information is usually not verified, it could exert a change in said perception through opinions. The intangible social construction that has been created within the use of social networks defines keywords about the collective experience. In this case, around numerous violent acts, where citizens manage to specify textually its broadest meaning in terms of their perception, since social networks are a space where human behavior and free will take place in real time, becoming a rather convenient to talk about concerns [12]. A study carried out in Colombia on the perception of insecurity focused on the emotional response that a person experiencing a crime situation could give, in this case the use of tools such as surveys did not dictate an appropriate convenience and focused their attention on the results provided for information based on social networks, specifically using Twitter [13].

Twitter is a social network where users share based on subjectivity, and transmit human habits by capturing and perceiving opinions. This information can be extracted from the Twitter API data access service to perform data analysis [14]. Another study was conducted in London using Twitter to explore and analyze patterns of reactions to homicides. This tool allowed to quantify the information of people indirectly affected by the events and based on the location, the speed of expansion of the news was analyzed [12].

Due to the need for a tool capable of measuring the effects of concern in citizens, this study proposes the extraction, and analysis of citizen perception variables using the Twitter social network as the information base. Daily tweets will be analyzed after 11 days, through filtering appropriate of data where key words of insecurity such as robbery and homicide are taken into account, which are contrasted through a survey of free expression on insecurity. These data are processed and analyzed statistically, at the national and regional levels with the most important

cities in Ecuador (Guayaquil and Quito). A correlation analysis was carried out at the regional level with the most important cities in the country, and high rates of crime in Guayaquil and corruption in Quito and Guayaquil generated by social networks were detected.

2. Materials and Methods

Initially, a survey of 200 respondents was conducted to determine the keywords that are highly influential in the colloquial language, with the aim of measuring their concept of insecurity. The compilation of the obtained words that are shown in Table 1 will determine the search criteria within the algorithm carried out in Python.

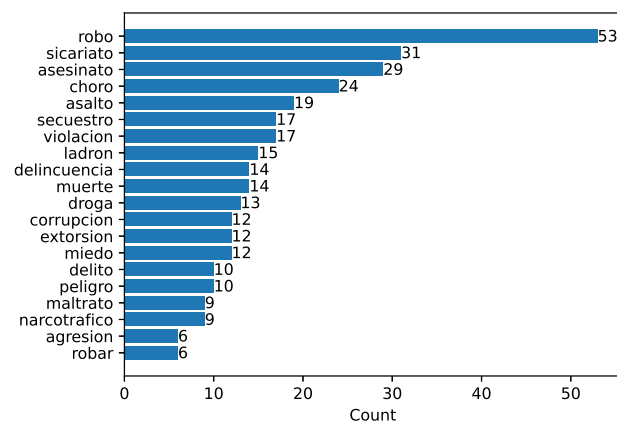
For the collection of the Twitter data, the words that were found most frequently within a data set resulting from the survey were evaluated as shown in Figure 2. Because the survey had an open text field to indicate the keywords, a tokenization process involving NLP (Natural Language Processing) techniques was carried out, in which each survey response passed through a text processing pipeline and lemmatization (obtaining its canonical form or lemma) using SpaCy library next to the module SpaCy Stanza [15], in this way we obtain the words with the highest incidence. As there is a constant variation in the format of the responses of the respondents, it was necessary to implement rules for tokenization through the Algorithm 1.

This procedure initially separates the words of each answer: by line if there are line breaks; by comma if there are commas; or by spaces if they only contain spaces. Once the responses

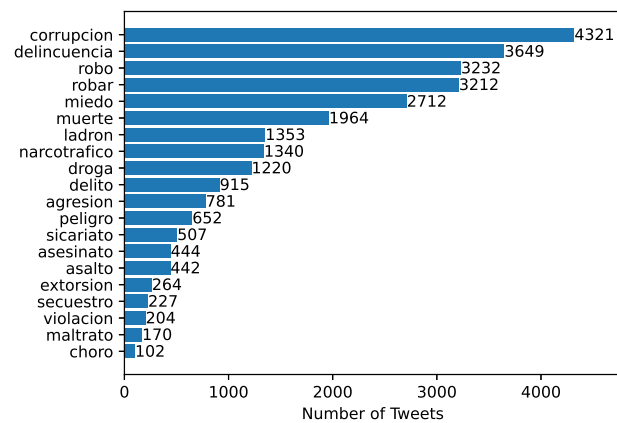
Table 1

Most frequent terms related to unsafety: Ecuadorian slang

| Survey Terms | Twitter Terms |
|--------------|---------------|
| robo | corrupción |
| asesinato | delincuencia |
| sicariato | robo |
| choro | robar |
| asalto | miedo |
| secuestro | muerte |
| violación | ladrón |
| ladrón | narcotráfico |
| delincuencia | droga |
| muerte | delito |
| droga | agresión |
| corrupción | peligro |
| extorsión | sicariato |
| miedo | asesinato |
| delito | asalto |
| peligro | extorsión |
| maltrato | secuestro |
| narcotráfico | violación |
| agresión | maltrato |
| robar | choro |



(a) Survey Report



(b) Twitter Report

Figure 2: Data collection related to insecurity words

are processed, the use of **Workers** is handled, which uses the "web" module of the Python Pattern [16] library, uses search filters and geographic location of the tweets. These workers allow parallel interaction with the Twitter API, maintaining the original structure of the tweets.

For the storage phase, the use of MongoDB was considered, which is a NoSQL (non-relational) database management system that stores data in the form of JSON (JavaScript Object Notation) documents instead of a columnar format [17]. This in-memory NoSQL database manager system was selected due to its speed compared to relational database managers, since it has been shown that its storage and insertion speeds are greater than relational systems [18].

The responses generated by the Twitter API are used as input to be stored in a **MongoDB** database. This data goes through a normalization process, separating the data into different collections (tables) by user, content of the tweet and metadata of the search carried out by the **Workers**. Before storing the tweets, here are subjected to a sanitization process, in which

Algorithm 1: Pipeline Steps

-
- Step 1 Input **Survey Data**;
 - Step 2 Separate words by available separator;
 - Step 3 Convert the words to lower case;
 - Step 4 Remove space and punctuation marks;
 - Step 5 Lemmatize words;
 - Step 6 Remove accents marks;
 - Step 7 Output **Tokens**;
-

symbols, labels, links and **#hashtags** are removed of the original content of the tweet and is stored in an additional field so as not to alter the word quantification process. The storage process prevents overwriting of data to be able to carry out a historical evaluation of each tweet. For this, a web service was developed that serves as a bridge between the collection workers and the database system MongoDB. Figure 3 presents the overview system architecture, along

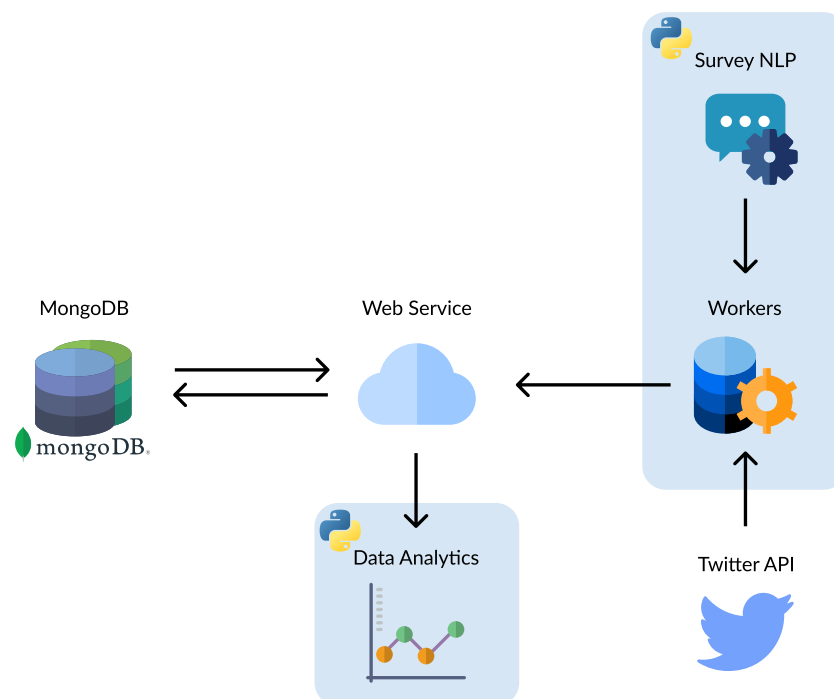


Figure 3: Architecture of the environment

with the data flow of each component. All components of the system have a flow unidirectional, except for the MongoDB database and the web service, which are capable of sending and receiving data each. For the analytical process of this research, used the web service as the data source for the analysis graphs, since it acts as a mirror of the Twitter API and has only a subset of relevant data.

3. Results and Analysis

To understand the levels of security that exist in a country and in its main cities; not only it is enough to have the data provided by the national government as a reference, but it is important to understand the relief of the people and public opinion too; since many of the events related to insecurity are mostly hidden due to fear of repression; especially in a Latin American context where culture and violence have a singular connotation. In Figure 1, which data is taken from the national government, they only report data on homicides and death in a six-month period; when insecurity has a deeper meaning; maybe it could be measured through the perception of feelings that generate panic or fear in citizens, and even an impact on the economic model of the productive matrix of a country. Taken based on the correlation of the words with the highest number of tweets in Figure 2, both for the survey report as well as for tweets, it has been explored in a weekly period due to the limitations of the Twitter API service, the number of tweets generated with the words with the highest incidence at the national and regional level as shown in Figure 4.

When we analyze the social and political crisis in Ecuador, people define robo, delincuencia, muerte, corrupción and miedo as the words that are most linked to insecurity. These curves will depend on the events with the greatest impact on social networks that may arise over time. On August 14 we have a report of approximately 600 tweets for a news item that caused panic in the city of Guayaquil [19], and this causes it to alter and generate commotion among citizens on social networks. This perception can be correlated to understand what happens with the most important cities in the national territory, where we have obtained some relevant data shown in Table 2.

We further observe that in Figure 4 (a) crime and corruption are the words of greatest concern; In other words, corruption is a factor that always affects our environment and can cause a perception of insecurity for investment in our national territory.

In Guayaquil, Figure 4 (b), being one of the most dangerous cities in the national territory,

Table 2

Correlation of terms related to unsafety

| Words | Ecuador | Quito | Guayaquil |
|--------------|---------|-------|-----------|
| robo | -0.619 | 0.544 | 0.108 |
| delincuencia | -0.715 | 0.171 | -0.305 |
| muerte | -0.578 | 0.202 | 0.017 |
| corrupción | 0.191 | 0.670 | 0.627 |
| miedo | -0.133 | 0.609 | 0.279 |

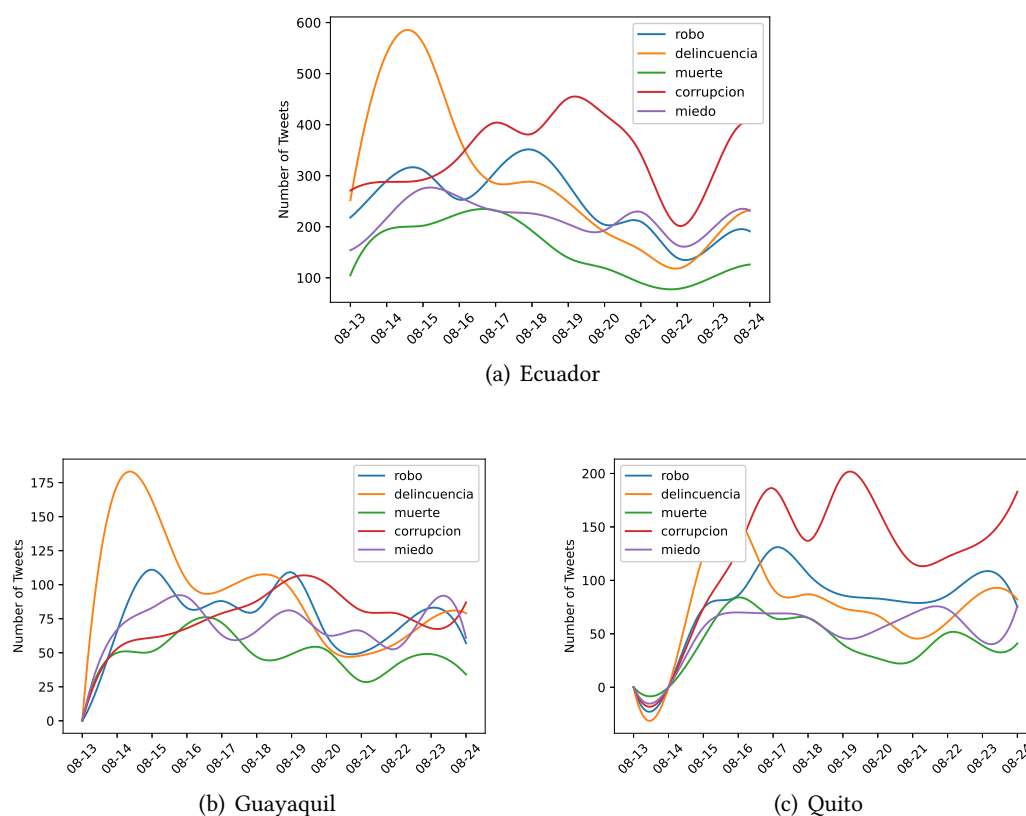


Figure 4: Data Analytics of Unsafety Perception

maintains uniformity among all the words that were selected over 10 days from August 14 to 24. In the city of Quito, the interaction of tweets is more marked with the focus on corruption as it is a city surrounded by politics, then robbery and below crime. An increase in the interaction of the tweets also coincides with the event raised on the date of August 14. This gives an approach to the fact that each city is a diverse reality due to issues related to cultural and urban connotation. On August 21 in the city of Quito, an event takes place that goes viral on social networks with **aggression** within a sporting event, where the increase in tweets is reported in the Figure 5, where clearly this word is not the most common in social networks, but requires a contrast in the proportional increase in the number of tweets for the words most related to security. This is affected from August 21 to 23 in the Figures 4 (a), (b), (c).

4. Conclusions

Through this study, the influence of real events on public perception and opinion in spaces of open discussion is analyzed from the content of social networks. In addition, the topics with the greatest social impact in terms of citizen insecurity in a given time have been identified

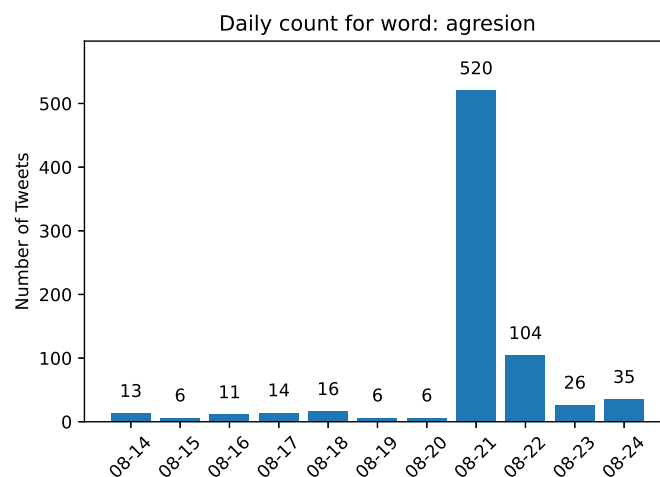


Figure 5: Analysis of aggression at national level

through the quantification technique, by extracting tweets.

This data can lead to a socioeconomic, geopolitical and productive study due to the cultural change of the masses. The increase in criminal events within the country leads to the need to implement public policies based on results of perception.

The statistical analysis carried out shows that there is a frequent fluctuation highly dependent on daily events. Likewise, these fluctuations can vary between the different subregions of the country from day to day. The most important correlation through this study was Ecuador - Guayaquil **crime** and **corruption** between Ecuador - Quito - Guayaquil. It was also observed that each event on social networks can have an impact on all the words linked to insecurity.

Predicting a crime rate is very complex since various criminal analysis had already confirmed that crimes are unequally distributed in place, time and context. Such kinds of situations are strongly driven by the environment, inequality, and lifestyle of Ecuadorian citizens, producing a high rate of victimization and negative effects for the people whose lifestyles forces them to expose themselves to a higher risk level.

The information that circulates in social networks responds to the perception and opinion of people on various topics. Regarding security, given the reality of indicators of violence and robbery in Ecuador, this issue is no exception and generates spaces for citizen opinion. In this research, the influence of networks on the perception of insecurity is verified, which contributes to the generation of a diagnosis that allows the construction of strategies for its control. It is necessary to highlight that this tool could allow the determination of an indicator, for the measurement of its evolution. The stochastic error that could be generated when unreal accounts unfoundedly seek to alter such perception in order to generate chaos or stability should also be taken into consideration prior to any conclusion, therefore the importance of reaching the largest possible filter so that the information processed is mostly from real accounts.

References

- [1] D. M. Boyd, N. B. Ellison, Social network sites: Definition, history, and scholarship, *Journal of computer-mediated Communication* 13 (2007) 210–230. doi:10.1111/j.1083-6101.2007.00393.x.
- [2] H. Gil de Zúñiga, N. Jung, S. Valenzuela, Social media use for news and individuals' social capital, civic engagement and political participation, *Journal of computer-mediated communication* 17 (2012) 319–336. doi:10.1111/j.1083-6101.2012.01574.x.
- [3] X. Wang, D. E. Brown, M. S. Gerber, Spatio-temporal modeling of criminal incidents using geographic, demographic, and twitter-derived information, in: *2012 IEEE International Conference on Intelligence and Security Informatics, IEEE, 2012*, pp. 36–41. doi:10.1109/ISI.2012.6284088.
- [4] C. S. Park, B. K. Kaye, Expanding visibility on twitter: Author and message characteristics and retweeting, *Social Media+ Society* 5 (2019) 1–10. doi:10.1177/2056305119834595.
- [5] A. Alvarado, La sociología del crimen y la violencia en américa latina. un campo fragmentado, *Tempo Social* 32 (2020) 67–107. doi:10.11606/0103-2070.ts.2020.175010.
- [6] J. Albarracín, N. Barnes, Criminal violence in latin america, *Latin American Research Review* 55 (2020) 397–406. doi:10.25222/larr.975.
- [7] M. E. Ávila, B. Martínez-Ferrer, A. Vera, A. Bahena, G. Musitu, Victimization, perception of insecurity, and changes in daily routines in mexico, *Revista de Saúde Pública* 50 (2016). doi:10.1590/S1518-8787.2016050006098.
- [8] I. D. Reid, S. Appleby-Arnold, N. Brockdorff, I. Jakovljević, S. Zdravković, Developing a model of perceptions of security and insecurity in the context of crime, *Psychiatry, psychology and law* 27 (2020) 620–636. doi:10.1080/13218719.2020.1742235.
- [9] K. M. Ortega, S. L. Pino, Impacto social y económico de los factores de riesgo que afectan la seguridad ciudadana en ecuador, *Espacios* 42 (2021) 52–70. doi:10.48082/espacios-a21v42n21p04.
- [10] M. Córdova Montúfar, Percepción de inseguridad: una aproximación transversal, *Ciudad Segura* 15 (2007) 4–9.
- [11] Instituto Nacional de Estadística y Censos, Encuesta de victimización y percepción de inseguridad 2011, <https://www.ecuadorencifras.gob.ec/encuesta-de-victimizacion-y-percepcion-de-inseguridad-2011/>, 2011.
- [12] O. Kounadi, T. J. Lampoltshammer, E. Groff, I. Sitko, M. Leitner, Exploring twitter to analyze the public's reaction patterns to recently reported homicides in london, *PLoS ONE* 10 (2015). doi:10.1371/journal.pone.0121848.
- [13] L. F. Chaparro, C. Pulido, J. Rudas, J. Victorino, A. M. Reyes, C. Estrada, L. A. Narvaez, F. Gómez, Quantifying perception of security through social media and its relationship with crime, *IEEE Access* 9 (2021) 139201–139213. doi:10.1109/ACCESS.2021.3114675.
- [14] L. M. Gómez, C. García Torres, Twitter, *Revista Colombiana de Anestesiología* 38 (2010) 539–540.
- [15] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, C. D. Manning, Stanza: A Python natural language processing toolkit for many human languages, in: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2020*, pp. 1–8.
- [16] T. De Smedt, W. Daelemans, Pattern for python, *J. Mach. Learn. Res.* 13 (2012) 2063–2067.

- [17] M. Polo-Usaola, MongoDB: gestión, administración y desarrollo de aplicaciones, Macario Polo Usaola, 2015.
- [18] F. Rubio, P. Vega, R. P. Reyes, Nosql vs. sql in big data management: An empirical study, *KnE Engineering* 5 (2020) 40–49. doi:10.18502/keg.v5i1.5917.
- [19] La Hora, Explosión en «cristo del consuelo», <https://www.lahora.com.ec/pais/explosion-en-cristo-del-consuelo/>, 2022.