Sentiment analysis in Twitter for the Medellin Development Plan

DS4A
CORRELATIONE ONE - COLOMBIA

FINAL REPORT - DATA SCIENCE PROJECT
JULY 10, 2022

GROUP 247
Team Members
Arteaga Juan, Caro Wilson, Quintero Jonathan, Luna Germán, Sosa Sergio, Vergara Gustavo

Contents

1 Introduction

Your team should explain how your solution is distinct from existing approaches to the problem and what value it adds over those.

1.1 Document description

To describe the proposed solution approach, several chapters are presented in this document that summarizes the most important stages of this project. **Chapter 3: Data Analysis and Computation**, describes the process of data access, download, and conformation of the datasets that will be the essential input for the realization of the project. In this chapter also there is the description of the exploratory data analysis in which the understanding of the data and a cleaning process are carried out based on different criteria that will be described in detail in order to finally obtain an adequate database for the model execution. Finally, this chapter presents the description of the chosen analytical model and its generalities and limitations and the obtained results according to the previous model. The document ends with a section of conclusions on the different previous stages, and the bibliographical references consulted for the execution of this project.

1.2 Who is this project for

2 Overview

2.1 Business problem and solution approach

The Medellin Development Plan (MDP) 2020 - 2023 [1] is the local government proposal that seeks to guarantee comprehensive attention to the needs of Medellin's citizens, care for vulnerable populations, economic reactivation, the construction of a sustainable city and the generation of opportunities based on a major educational transformation.

In other words, it is a promise of public policies by the mayor's office in order to improve the welfare of the people they represent by trying to meet the needs that afflict them. But are these in tune with the proposals presented in the city's development plan?. This is what this project will try to answer, for which it will divide the analysis into two components:

1. Is the MDP aligned with the expressed by the population?
    - In this first component, the main problems expressed by the inhabitants of Medellin prior to the entry into force of the MDP will be identified and it will check if these are part of the development plan.
    - To solve this question a text analysis of information extracted will be performed from Twitter (needs) and MDP (proposals) to check if there is any relationship between them.
2. What is the perception of citizens in relation to MDP programs and projects during the government's term?
    - The response of the citizens of Medellin in relation to this topic will be analyzed through a sentiment analysis based on the perception exposed by the inhabitants on Twitter in relation to these projects and programs of the MDP.

2.2 Business Impact

Through the citizen's perception of the MDP programs, as well as the needs felt by the population on Twitter, it is possible to establish an alternative measurement of the assertiveness of the current local government plan and likewise to open up the possibility of configuring a more appropriate action horizon for future development plans for Medellin aligned with the above.

2.3 Application Overview

This should cover what the application does, what the primary use cases are, and how a user would interact with it.

3 Data Engineering

3.1 Interactive Front-end

3.2 Database

4 Data Analysis and Computation

4.1 Dataset

To get the dataset from Twitter, it was necessary to create a developer account in this social network to access the developer platform. Then, by this account, it was possible to access the creation of a project and to request credentials to be able to do queries from Twitter by its API. The Twitter API enables programmatic access to Twitter in unique and advanced ways, it taking advantage of the core elements of Twitter like: Tweets, Direct Messages, Spaces, Lists, users and more. The API of Twitter has three access levels: Essential, Elevated, and Academic Research.

For the consolidation of the project's data set, some credential requests were made to query and download the tweets using the Twitter API. Initially it was possible to obtain an Elevated credential that allowed a historical time window of 7 days before the consultation date. By using this credential it was possible to perform some queries and obtain a first and preliminary set of data. This data set was consolidated for 13 days (May 9 - May 21), due to the historical time window limit granted by the credential to obtain information. Through these first queries it was possible to understand how the API worked and determine the variables to use in the project.

However, after a second request to Twitter, it was obtained a Researcher credential to use Twitter API. Through the use of these credentials, it was possible to get access to even more data and advanced search endpoints. Through final researcher credentials, the project datasets were finally consolidated.

4.1.1 Query - API Twitter

The Twitter API allows you to perform a variety of different actions using code. So the first step was to establish the credentials and connect to the Twitter API. "Credentials" refers to those access keys to the Twitter developer account. Once the connection with the Twitter API was made, it was possible to use the available query options. For this project, the tweet search option was used, defining the following search terms: predefined keywords contained in the tweets, the start and end date points of the search, and a list with the variables required to obtain such as the text of the Tweet, the location, the number of retweets, between others. These selected variables that conform the dataset will be explained in Section 2.2.

Then, some Python code functions were created to perform and store the information properly. This process is presented in steps as follows:

1. Download recent tweets as a list. Filters have been made to delimit tweets related to Medellin and predefined keywords.
2. Convert the list downloaded by the previous function into a Dataframe.
3. Modify structure of the text, allowing to save the tweets without generating any conflict with the database.

4.1.2 Tweets 2019

To carry out the present project, a first query was performed search for the 2019 year in order to get the most popular needs and concerns about Medellin expressed on Twitter in a previous window time to the MDP term (2020 - 2023).

Creation

This dataset was created by the search query in the API, which was performed by putting "Medellin" as a parameter in the search, and 01 January to 31 December of the 2019 year as the beginning and end date points. The file size of this dataset was 4.3 MB and contained a total of 17700 tweets.

Contents

The dataset variables selected and stored such as columns are presented in Table **4-1**. In this table, it is also presented a brief description of the variables and their datatype.

No	column name (variable)	datatype	description
1	full text	string	full text of the tweet
2	user	string	username who posted the tweet
3	location	string	location where the tweet was post
4	date	datetime	time when the tweet was post
5	tweet id	int	primary key, number id of the tweet
6	number rt	int	number of retweets of the tweet
7	number likes	int	number of likes of the tweet
8	number reply	int	number of likes in the reply
9	conversation id	int	identification number of conversation

Table 4-1: Summary description of dataset variables.

Purpose

This dataset was created to make a first exploratory analysis in order to obtain a first approach and data understanding about the most relevant issues and concerns of the population about Medellin on Twitter. This was done to verify if these issues are correlated or included in the projects and strategic lines of work proposed in the MPD.

4.1.3 Tweets keywords 2019-2022

Creation

This second dataset was created after the data understanding was performed, this stage included the exploratory data analysis of the Tweets 2019 and the MDP document. In order to consolidate this first dataset, denominated in this project such as **Tweets keywords 2019 - 2022**, various search queries were done in the API for the years 2019, 2020, 2021, and so far 2022 taking search parameters a list of strategic keywords extracted from the MDP document, oriented to the lines of the plan to be measured. The process of selection of these keywords will be present in the Section 3.3.

Initially were consolidated a dataset by keyword and year. Then, a function was coded in order to consolidated all keywords-year dataset in just one that corresponds to this dataset. The data was stored in a CSV file for later description and exploration. The file size of this dataset was 90.9 MB and contained a total of 303,008 tweets.

Contents

This dataset contains the same variables presented in Table **4-1** for 2019, 2020, 2021 and so far 2022, which was consolidated by tweet queries for these years taking as search parameters (Table **4-2**) the extracted keywords list of the chosen strategic lines to analyze in this project of the MDP.

column name	datatype	description
id key word	int	primary key
key word	string	key word used to search the tweet

Table 4-2: Summary description of keywords search.

Purpose

This dataset was created in order to make the exploratory data analysis for the **Tweets 2019 - 2022**.

4.1.4 Dataset Model

Creation

This dataset was created as a result of the data preparation stage on the **Tweets keywords 2019 - 2022** after a cleaning process. This data preparation stage through this cleaning process will be presented in Section 3.2.

Contents

The distribution of the data is the same as **Tweets keywords 2019 - 2022** in terms of variables, but contains just the necessary information in the Tweets text and other variables to run out the model.

No	column name (variable)	datatype	description
1	full text	string	full text of the tweet
2	user	string	username who posted the tweet
3	location	string	location where the tweet was post
4	date	datetime	time when the tweet was post
5	tweet id	int	primary key, number id of the tweet
6	number rt	int	number of retweets of the tweet
7	number likes	int	number of likes of the tweet
8	number reply	int	number of likes in the reply
9	conversation id	int	identification number of conversation
10	id key word	int	primary key
11	key word	string	key word used to search the tweet

Table 4-3: Summary description of Model dataset variables.

Purpose

This dataset was conformed in order to get suitable data to carry out the sentiments analysis by the chosen model.

4.2 Exploratory Data Analisis - EDA

4.2.1 Data understanding I

EDA: Tweets 2019

The first problem project is focused on meeting the needs of the population of Medellín prior to the validity of the MDP (2020-2023). This may seem impossible or it could have bias issues since it only will work with data from one social network. Therefore, a proxy variable will be used that allows knowing the relevant issues and the concerns expressed by people about Medellín.

To start with the exploratory data analysis, the dataset called **Tweets 2019** will be used, which contains the variables presented in Table **4-1**. The objective of this analysis is to know the dynamics of interaction on the Twitter platform, and the usefulness of the queries made to form this data set.

The first variable of interest will be the text of the tweet, this variable will give information about the composition of the tweet and also some inputs to improve the API request to excel in the sentiment analysis model.

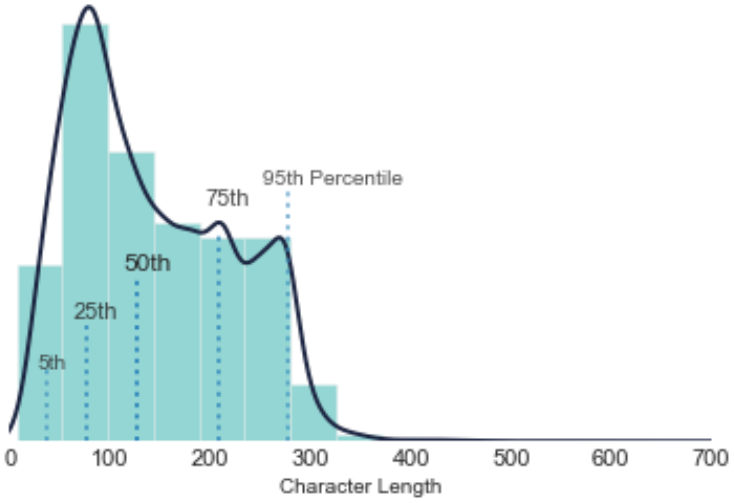

Figure 4-1: Character tweets distribution by length.

Figure **4-1** describes the length of text tweets measured in characters, with a mean of 143 characters, a maximum of 689, and a minimum of 7 characters, which shows a right-skewed distribution, so there are outliers that could reflect controversial issues or just noises. However, it is necessary to continue with the exploratory data analysis.
The behavior of Figure **4-2** distribution provides additional information, with a mean of 22, a maximum of 75, and a minimum of 1 character. This reinforces the same behavior of Figure **4-1**,

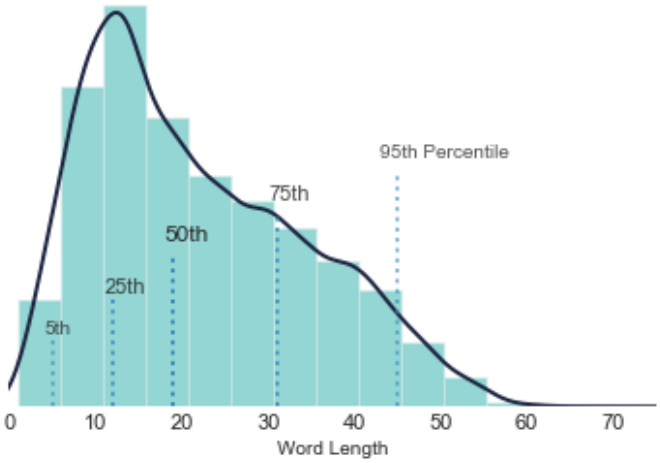

Figure 4-2: Word tweets by length.

a biased right distribution in which the outliers are responsible for this fact.

Separately, both figures do not provide additional information, but when comparing them, some key pieces come into the picture. Focusing on the 75th percentile for Figure **4-1**, it finds a distribution with sharp spikes, unlike the distribution of Figure **4-2** in which there is a smooth behavior. A proposed hypothesis to explain this could be the values of this distribution by character length through the use of emojis, punctuation marks, mentions, hashtags, white spaces or words like "hellooo". The presence of these type of elements in the text of the tweet distances its length value from the central tendencies.

For the project purpose, **"word"** is defined as a characters sequence separated by a single white space. For this reason, the Figure **4-2** distribution presents a shape with less sharp points in the third quartile, from this "Hello @Medellin 😄🤗👋" (20 characters) to this: "Hellooo|@Medellin|😄🤗👋" (3 words). Knowing the length distribution of tweets text provides an approximation of how much relevant information is found in them, however, it is a partial analysis because its content is unknown.

Table **4-4** shows a sample of 10 tweets with two words or less. In this case, the most relevant word is "Medellín" along with emojis and mentions, among others. This fact was expected because this word was the main search parameter for the API query and the main subject of analysis.

Figures **4-3** and **4-4** show some particular tweets with the highest number of words. These figures highlight a large number of mentions and also the fact that the word "Medellín" depends on the context. For example, "Medellín" could refer to the Colombian professional soccer league,

Full_text	Num_words
“@from_oz Medellin”	2
“@SaqueLargoWin Medellin”	2
“@Ricardo_Arjona Medellin”	2
“@BillieEilishCOL Medellin”	2
“@SaqueLargoWin Medellin”	2
“Medellin 😍”	2
“Medellin!!!! 😘”	2
“Medellin, Extremadura”	2
“Medellin”.	1
“😘 Medellín”	2

Table 4-4: Sample of full text tweets with two words or less.

in relation to ”Deportivo Independiente Medellín”.

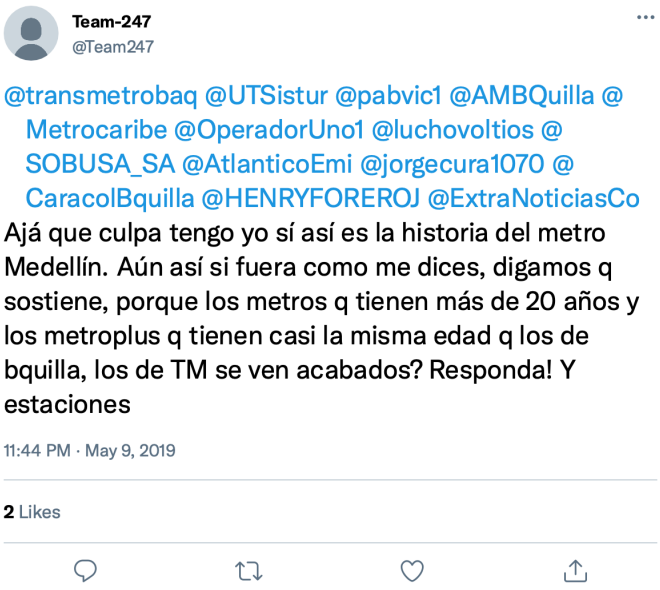

Figure 4-3: Tweet with the highest number of words (75 words).

Another factor to identify the tweet’s importance is its number of likes, retweets, and replies. This could be another proxy variable in order to know the problematical in Medellin because it is expected that these topics are the most interactive. Table **4-5** summarizes the tweets with the highest number of retweets/replies, likes, and second-highest retweets.

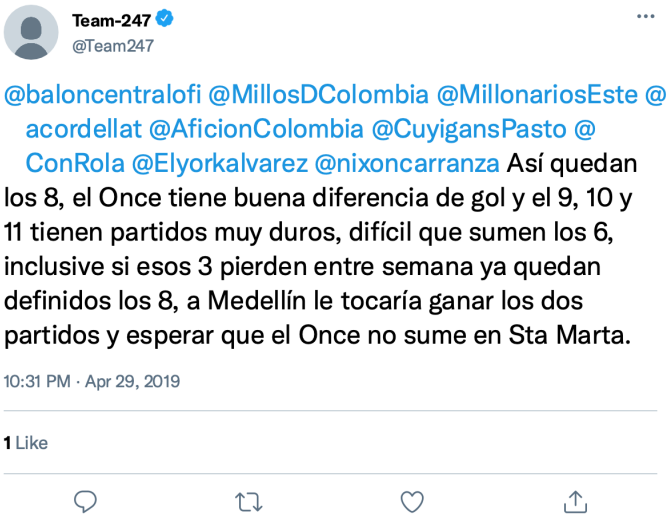

Figure 4-4: Fourth tweet with the highest number of words (65 words).

Description	Full text
Maximum number of retweets/replies	"Miren la diferencia entre una estación de Transmilenio Bogotá y otra de Metroplus (el Transmilenio de Medellín). El problema no es el tamaño de la estación... ....es cultura"
Maximum number of likes	"Alejandro Gómez @AlejandroGL2014 es médico y especialista en salud y finanzas de EAFIT. Fue director de Salud Pública en Medellín, gerente de la red hospitalaria de esa ciudad, y Director Nacional de Nutrición del ICBF y será el nuevo Secretario de Salud de Bogotá. #BogotáCambia"
Second highest number of retweets	"La falta que nos ha hecho en Bogotá tener el sentido de política pública de Medellín es incuestionable."

Table 4-5: Retweets, replies and likes analysis.

From this analysis it is identified that "Medellín" does not only refer to a geographical location. This word is also used to make comparisons or comments using this word as a reference city. Some main topics of the city such as "Metro" appear in the tweets and could be related to some of the needs of the people of Medellín.

Likewise, it is to be expected that the majority of the tweets are located in the city of Medellín because the search was carried out with this word. Figure **4-5** presents a frequency bar chart visualization by the source location of the tweets. Nevertheless, scattered locations were found that also refer to the target city. As an example, a tweet from the city of Arequipa, Peru was

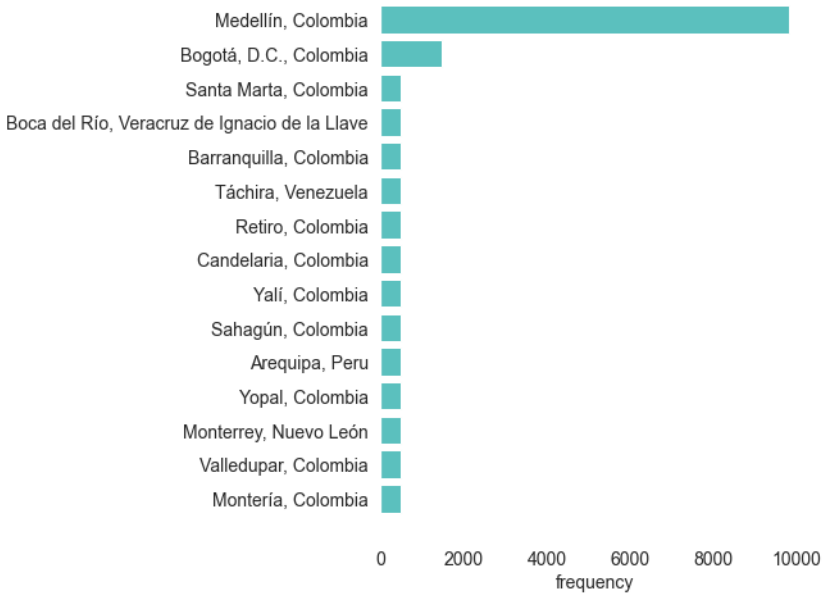

Figure 4-5: Tweets by source location.

extracted and this was the result: "Ahora que acaban de meternos 3 pepas, con Guerrero incluíso, voy a destapar un Medellín y prepararme un cuba mientras escucho a Janis. Que tengan buenas noches". A funny thing is that when referring to Medellin, it does not refer to the city but to the Ron Medellin liquor. This only confirms that a word without context can lead to unexpected results.

Figure 4-6: Word cloud full text Tweets 2019.

When graphing the frequency of words within the text of the tweets (Figure **4-7**), those that top the list are: pronouns, prepositions, definite determiners, quantifier determiners, punctuation marks and the word "Medellín" with its variants. These types of words are intrinsic to the use of language and hence their high recurrence. The word "Medellín" responds to the search filter

used in the API request, for our purpose, the meaning of these words would not provide more information to classify them in any sentiment.

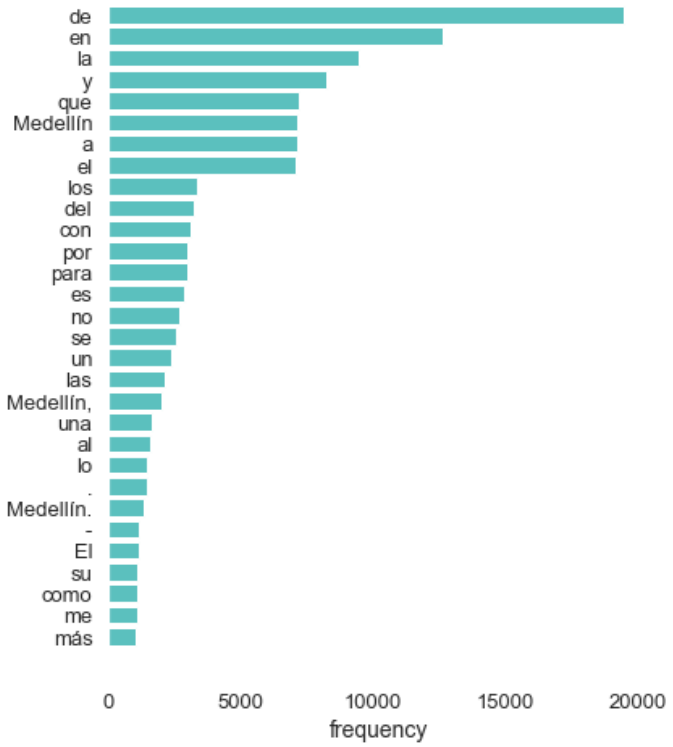

Figure 4-7: Words frequency Tweets 2019.

From the above, the following conclusions are obtained:

1. The large use of mentions with the at-word extends the length of the tweet and would not provide relevant information for sentiment analysis so its elimination could be considered.

2. It should be taken into account that the word "Medellín" depends on the context, i.e., it will not always refer to the city, but also to the soccer team, to the alcoholic beverage, as a comparison, or to other meaning. So it has to be careful when drawing conclusions and narrowing your search.

3. When finding different places it is to be expected that any resident abroad can mention the word "Medellin" and its different connotations, so it may not reflect problems in the city. When interpreting the results, grouping by localities or focusing only on the city of Medellin could solve this limitation.

4. The words found most frequently in the text of the tweets do not seem to be determinant for a subsequent classification model. Their omission would facilitate the processing of

the text that will serve as input. This stage will be expanded in the text pre-processing section.

The EDA of this section is only a first approximation to the complex process to extract relevant information from a social network. People express themselves in very different ways and the different grammatical rules cloud any indication of words that allow them to be associated to the target problem. Therefore, it is necessary to clean the text of the tweets to go further in the analysis. This entire process is summarized in Table **4-6** in the following section.

4.2.2 Data preparation I

Text pre-processing Tweets 2019

The original text of the tweets is subject to mentions, misspellings, emoticons, punctuation marks, jargon or words indigenous to the regions, i.e. a mixture of many elements. Therefore, a step-by-step cleaning process is presented to facilitate its handling. First, the unification of upper and lower case letters in the text is performed, converting them into lower case letters, so that words that are written the same but use different cases are identified as the same sequence of characters. For example, any user who types "Medellin" or "medellin" means the same thing, but the programming language recognizes them as different character sequences.

Then, the process of cleaning up the mentions and hashtags within the text of the tweet is performed. The mentions in the tweet describe user names or accounts, these are usually numerous and do not follow established patterns, so each of these is unique in its manipulation. However, these elements could be used in a later analysis to associate particular accounts or profiles that can be used in negative or positive comparisons in order to extract sentiments, but at this stage they do not provide any contribution to identify the needs of the population of Medellín.

Hashtags are usually associated with social trends, so their use may be linked to a specific period of time, this fact could bias the analysis to be performed with respect to the data sample that has been obtained. An example of this could be the following hashtag: "*#ParoNacional21Nov*", this was a trend after the social demonstrations occurred in the country for the year 2021, if the dataset had been extracted only for this year, such trend would have had a greater relevance or significance to identify the needs of the population in the analysis. Likewise, the topics or words used within the trends could be redundant as they may already be included within the text of the tweet.

Then, we proceed with the cleaning of punctuation marks, special characters and accents, since these types of characters do not provide relevant information that can be used in the analysis. On the one hand, punctuation marks, as seen in section 3.1.1 (**EDA - Tweets 2019**), are often used excessively and on the other hand, accents present the same problem mentioned about

Description	Tweet text
Raw text	“ Inteligencia artificial, Internet de las cosas y #blockchain son las tecnologías en las que se va a enfocar el Centro para la #4Revolución de @wef en #Medellín -@AlcaldiadeMed. @MincomercioCo @Ruta_N #EconomíaNaranja (vía @larepublica_co) ”
Unify the case of the text: convert it to lowercase.	“ inteligencia artificial, internet de las cosas y #blockchain son las tecnologías en las que se va a enfocar el centro para la #4revolución de @wef en #medellín -@alcaldiademed. @mincomercioCo @ruta_n #economíanaranja (vía @larepublica_co) ”
Remove the mentions	“inteligencia artificial, internet de las cosas y #blockchain son las tecnologías en las que se va a enfocar el centro para la #4revolución de en #medellín -. #economíanaranja (vía)”
Remove the hashtags	“ inteligencia artificial, internet de las cosas y son las tecnologías en las que se va a enfocar el centro para la de en -. (vía)”
Remove punctuation marks and special characters	“ inteligencia artificial internet de las cosas y son las tecnologías en las que se va a enfocar el centro para la de en vía”
Remove accents	“ inteligencia artificial internet de las cosas y son las tecnologias en las que se va a enfocar el centro para la de en via”
Remove stopwords	“ inteligencia artificial internet cosas tecnologias enfocar centro via ”
Remove whitespaces at the beginning and end of the text	“inteligencia artificial internet cosas tecnologias enfocar centro via”

Table 4-6: Text tweets pre-processing.

the text of the tweet, the encoding of each one is different, i.e. an accent or tilde omitted in a word results in a different set of characters when it is intended to refer to the same thing.

Finally, when finding the words with the highest frequency within the text of the tweet, it was observed that the categories presented in Figure **4-7** were important, and therefore it was difficult to see those words that reflected feelings or needs in an explicit way. Therefore, a list of *stopwords* was built as shown in the word cloud (Figure **4-8**) which were extracted from the text. In a last step, the spaces at the end and beginning of the tweet were eliminated.

After performing the cleaning process described in Table **4-6** a word cloud with the text of the tweets was obtained (Figure **4-9**). According to this figure, the relevant topics or with the high-

Figure 4-8: WordCloud - Tweets 2019 Stopwords .

est interaction in the Tweets - 2019 dataset can be observed. The top 5 words with the highest interaction were: “ciudad” (787), “via” (699), “metro” (435), “alcalde” (411) and “nacional” (399).

Figure 4-9: WordCloud - Tweets after text pre-processing.

Analyzing the results of Figure **4-9**, it was identified that not all the most frequent words referred to topics of interest for the city of Medellin. Therefore, making a more rigorous analysis, it was determined a list of categories (denominated in this project ***keywords***) that encompassed transcendent topics to which the tweets alluded. Figure **4-10** shows a bar chart with the incidence of these words within the dataset **Tweets - 2019**.

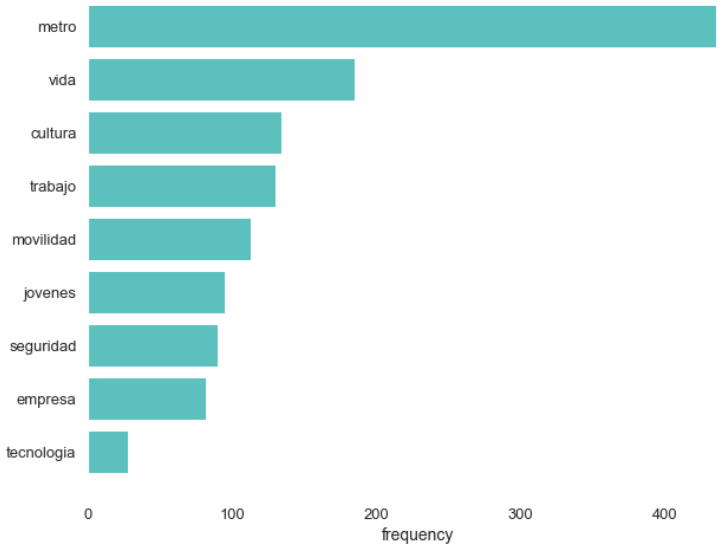

Figure 4-10: Barchar - Keywords.

4.2.3 MPD Text Analysis

The purpose of the MPD document is to establish the different guidelines and projects in terms of public policies for the city of Medellin in the period 2020-2023. In order to answer the first question proposed in section 1, a text analysis was carried out on the MPD with the purpose of determine if the topics expressed in this document were aligned with the categories obtained in the previous section. For this analysis, the same cleaning process used in Section 3.2.1 was carried out to obtain a suitable set of words with which to perform the required comparison.

Figure 4-11: Word cloud - MPD Text.

As a result of this process was obtained a word cloud (Figure **4-11**) from the PMD text. From this Figure, it is possible to observe words that describe a project execution and the metrics to evaluate its impact and evolution. Another relevant fact is the presence of words such as "seguridad", "vida", and "cultura", which are included in the list of ***keywords*** in section 3.2.1. After making this first approximation, the incidences of the keywords within the PDM text were searched, obtaining the result of Figure **4-12**.

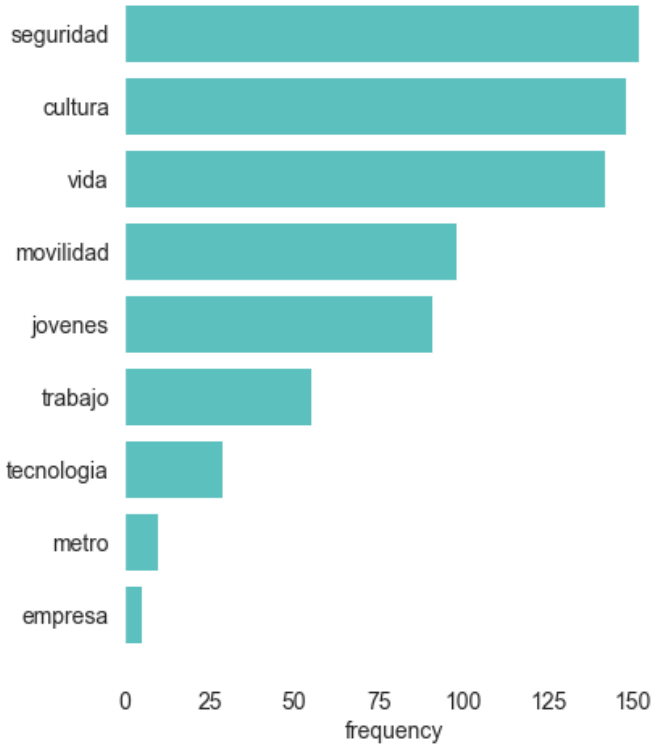

Figure 4-12: Keywords frequency in PDM text.

Within PDM document there is a chapter called "Líneas Estratégicas" which establishes the different strategic lines of action that encompass the proposals of the government plan and its execution in relation to the most important issues for the city future. For this reason, it was performed a text analysis in order to obtain the most relevant words (the words with higher frequency in the text) for each PDM strategic line (Figures **4-13** to **4-16**).

The first strategic line (wordcloud Figure **4-13**) seeks to create a digital culture and economic reactivation that will improve the quality of life of the population of Medellin through the management of new opportunities, education, entrepreneurship and job creation in areas associated with the digital economy and the fourth industrial revolution. This objective is closely associated with the words "life", "culture" and "technology".

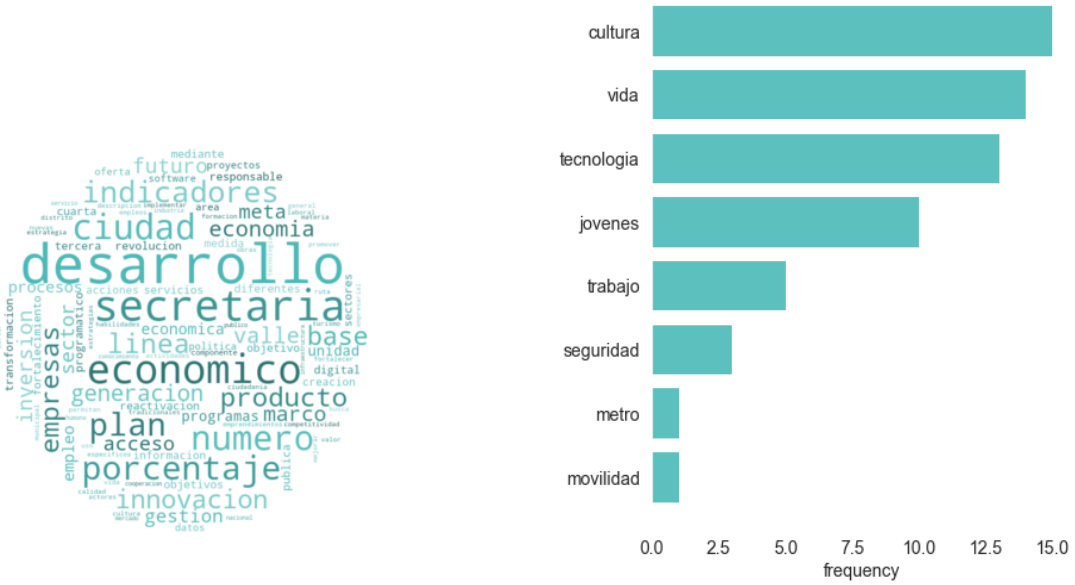

Figure 4-13: Word Cloud - "Linea estrategica 1: Reactivación Económica y Valle del Software".

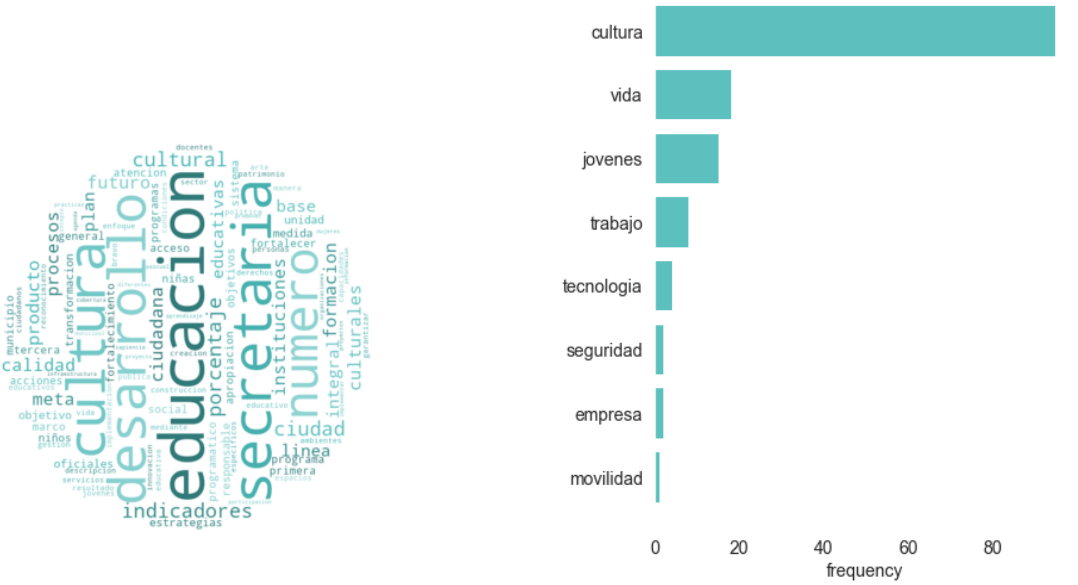

Figure 4-14: Word Cloud -"Linea estrategica 2: Transformación Educativa y Cultural".

The second strategic line (Wordcloud Figure **4-14**) seeks to articulate the city with cultural projects that strengthen the creative potential of citizens, safeguarding their heritage and memories, making Medellin a more supportive and peaceful city. It also contains programs focused on youth, gender equality, education, arts, and science. This makes it easy to explain the frequency of the keywords.

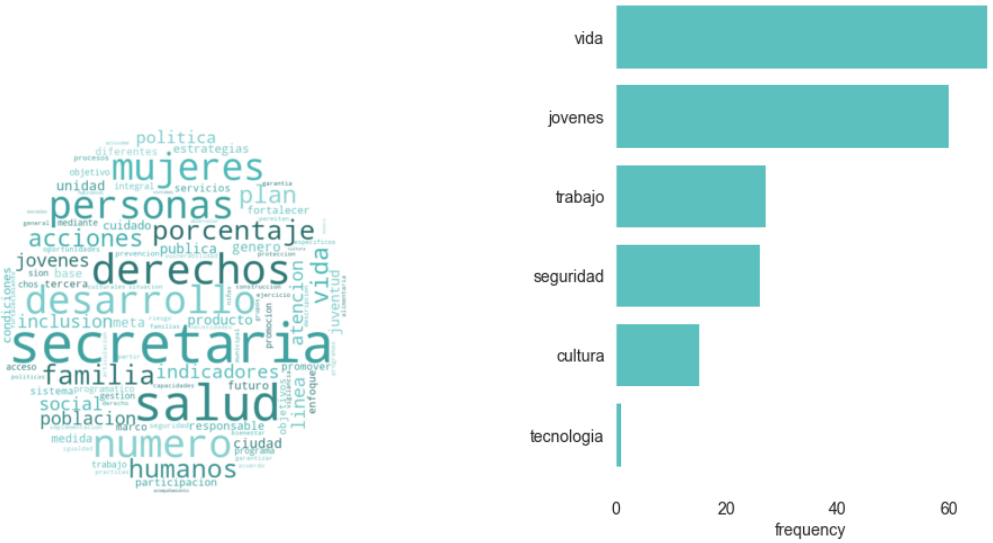

Figure 4-15: Word Cloud -"Linea estrategica 3: Medellín Me Cuida".

Line strategy 3 (wordcloud Figure **4-15**) focuses on the citizens of Medellin, in promoting, creating and guaranteeing basic and cultural living conditions, in order to have the ability to enhance their human and individual talents. Likewise, it promotes the generation of social, community, healthy, creative, safe and sustainable environments. In addition, in this line there are programs established for youth from public health to the so-called "Valle del Software".

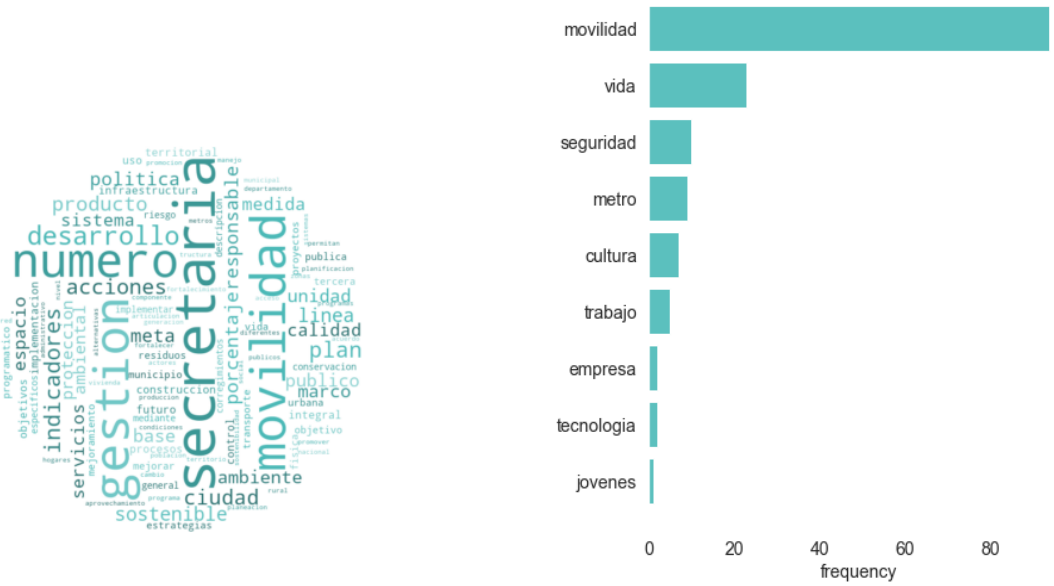

Figure 4-16: Word Cloud -"Linea estrategica 4: Ecociudad".

The fourth strategic line (Wordcloud Figure **4-16**) seeks to move Medellin towards a future of

sustainability, in which dignified habitability is guaranteed for its inhabitants and functional and harmonious integration with rural areas. In this line, programs for sustainable and intelligent mobility are highlighted in which include clean energies and cultural transformations, focusing on the conservation of all forms of life.

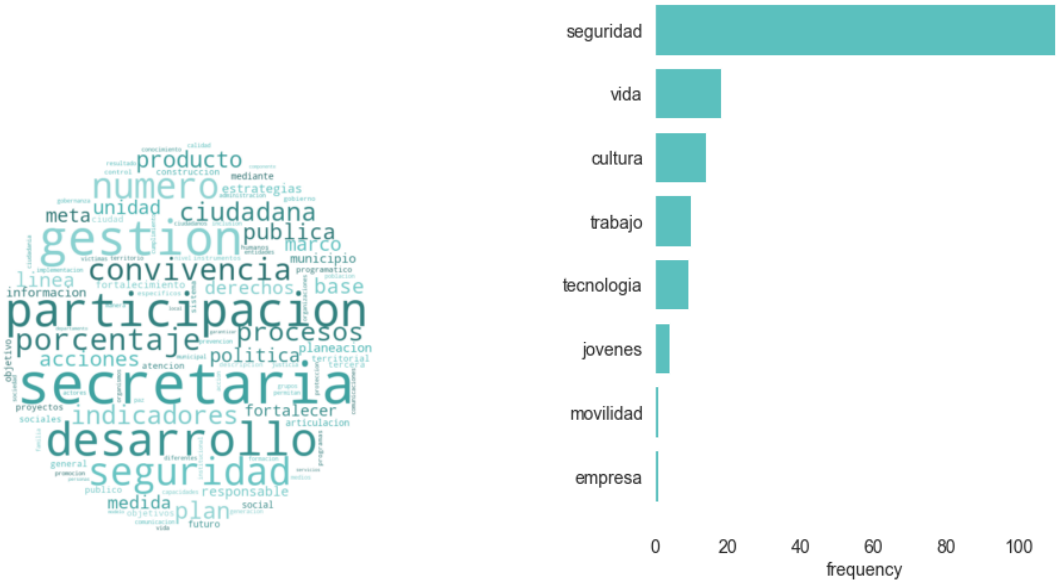

Figure 4-17: Word Cloud -"Linea estrategica 5: Gobernanza y Gobernabilidad".

The last line strategy (Wordcloud Figure **4-17**) aims to reinforce the synergy between government and citizenship, an open dialogue from the different knowledge, the consensus between the different actors and the collective construction of the citizen territorial peace process. Victims and justice are one of the components of this line, as well as security in terms of citizen coexistence and cybersecurity.

4.2.4 Data understanding II

EDA - Tweets keywords 2019-2022

Figure **4-18** describes the quantity of tweet by weekday. It is observed that the days in which there is more interaction of Tweets are located on weekdays, with a significant reduction of tweets on weekends. Figure **4-19** presents the quantity of tweets in the dataset by year, the year 2019 stands out with the highest proportion of tweets, followed by 2020,2021 and 2022 with the lowest quantity since it was only possible to extract tweets for 6 months (January-June). Finally, for 2021, it is observed that June was the month with most tweets. This can be explained by the fact that it coincides with the previous date for the second round of the presidential elections.

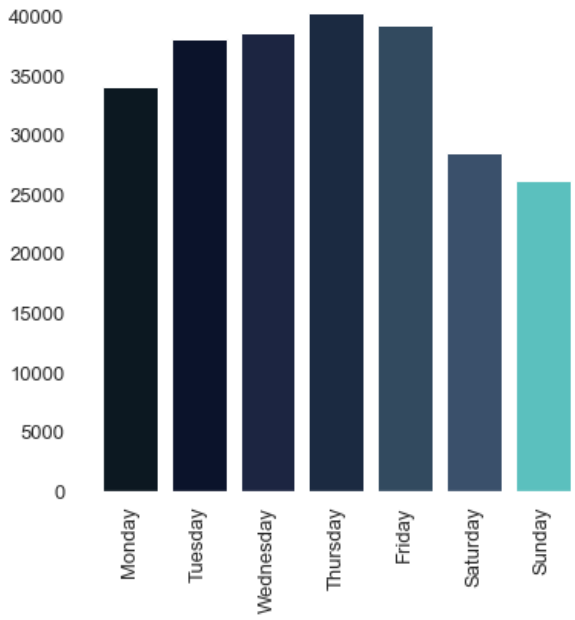

Figure 4-18: Tweets frequency by weekday.

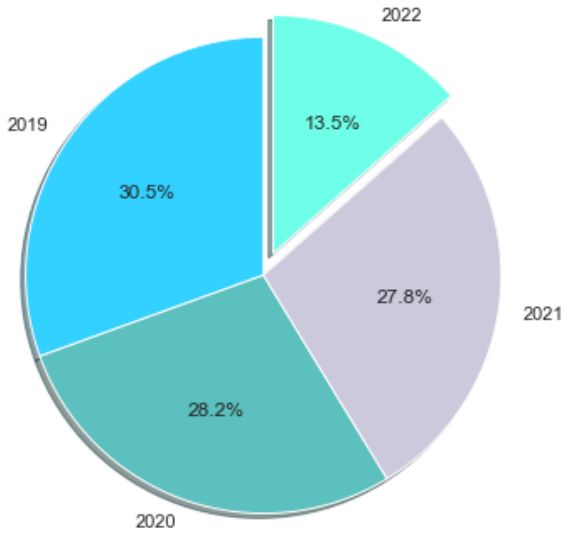

Figure 4-19: Tweets distribution by year.

In Figure **4-20** it is presented the tweets frequency by year. For 2019 the month with the highest number of tweets was October, the month in which the regional and municipal elections were held. In 2020 May was the month with most interactions.

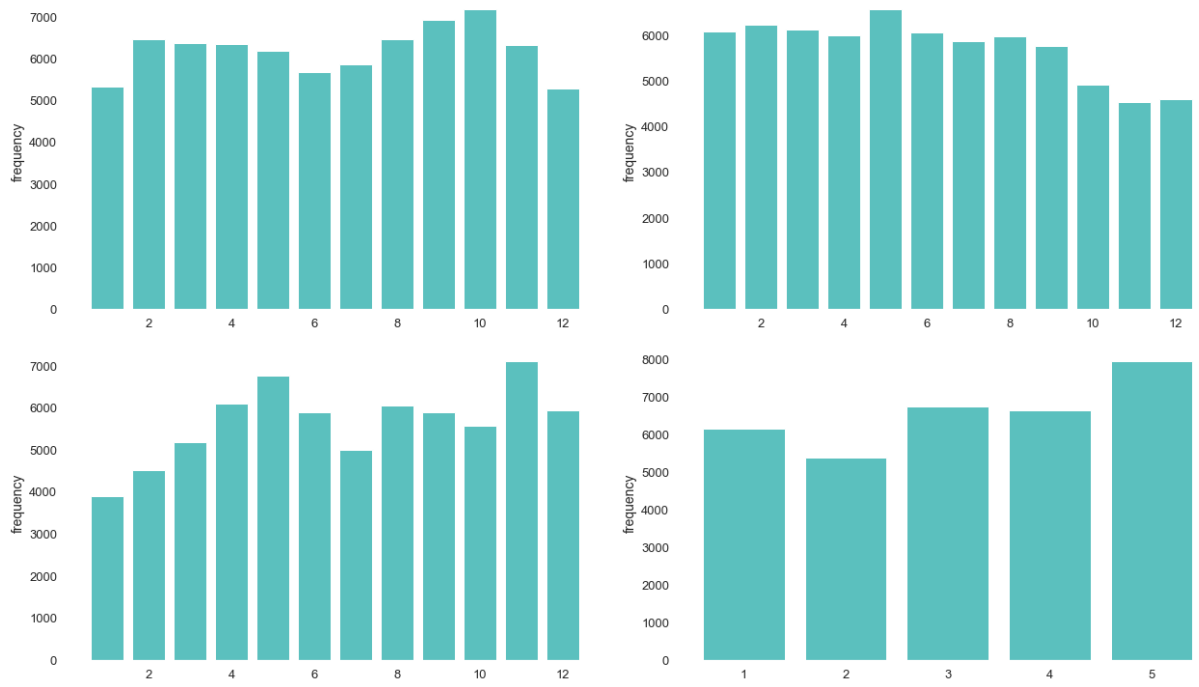

Figure 4-20: Tweets frequency for each year by months.

4.2.5 Data preparation II

Text pre-processing Tweets 2019-2022

To obtain the dataset for the model, the same cleaning process presented in Table **4-6** was carried out as the one presented for the 2019 dataset, in this case for the 2019-2022.

4.3 Statistical Analysis and Machine Learning

4.3.1 Model

5 Conclusions and Future Work

6 Bibliography

[1] Alcaldía de Medellín: Plan de desarrollo Medellín futuro 2020 - 2023. 3 (2020), Nr. 1, 1–1543. http://doi.wiley.com/10.1111/ens.12293%0Ahttp://dx.doi.org/10.1016/j.meegid.2010.01.004%0Ahttp://dx.doi.org/10.1016/B978-0-12-384890-1.00016-9%0Ahttp://dx.doi.org/10.1016/j.forsciint.2010.12.004%0Ahttps://doi.org/10.1016/j.forsciint.2018.12.002%0Ahttp://d. – ISBN 9788578110796