FCTNLP: Fighting cyberterrorism with natural language processing

Autor
Andrés Felipe Zapata Rozo

Trabajo presentado como requisito para optar por el título de Profesional en Matemáticas Aplicadas y Ciencias de la Computación

Director
Daniel Orlando Díaz López

Escuela de Ingeniería Ciencia y Tecnología
Matemáticas Aplicadas y Ciencias de la Computación
Universidad del Rosario

Bogotá - Colombia
2021

Índice

1. Abstract **2**

2. Introduction and motivation **2**

3. Objectives **3**

4. Methodology **3**

4.1. First Semester . . . 3

4.2. Second Semester . . . 4

5. Results **4**

5.1. First Semester . . . 4

5.1.1. Natural Language Processing review . . . 4

5.1.2. Profiling Twitter User Accounts . . . 5

5.1.3. Initial Architecture . . . 6

5.1.4. Initial architecture test . . . 7

5.2. Second Semester . . . 8

5.2.1. Extended experiments . . . 8

5.3. Final Architecture . . . 8

5.3.1. Documentation and presentation of the results . . . 9

6. Conclusion and Future Work **9**

Annex **10**

1. Abstract

The social networks are a rich source of data and have been used to promote or organize cybercrimes that affect the real world. Because of this, the law enforcement agency are interest in the crucial information that can be get on this sources. The amount of information and the informal language which is used to spread information makes the Natural Language Processing (NLP) and excellent tool to make analysis over post in social media. That is why, in this proposal an architecture with three NLP models are integrated to provide an exhaustive analysis from open sources like social media. This analysis extract entities from the text, identifies clusters of users and their respective polarity, finally all of the results are related in a graph database. This architecture was under test using data from a real scenario in order to determine their feasibility.

2. Introduction and motivation

The fight and prevention against crime is present in all the world and is the principal objective of the Law Enforcement Agencies (LEAs). With the context of telecommunication technologies new ways to commit crimes were appear. The acts that uses computer technology to commit a crime are consider cybercrimes [1]. Due to this, the new cyber domain began to be consider [4].

One of the cybercrimes that cause the greatest concern in global community is the cyberterrorism [8]. The terrorism and consequently the cyberterrorism are crimes difficult to define as many authors recognized, in this case we gone to consider the most accepted definition that is the use of terror as a weapon to intimidated the population for political, social or religious purposes [5, 6].

Social media are protagonists in people's daily lives, are platforms that provide an easy and direct access to a huge amount of content. Despite of the amount of the data available to their users the variability of the content that a user consume is low. This phenomena is called echo chamber and promotes the polarization between different clusters of users [2]. This polarization can be dangerous to the entire community, and example of this in the context of vaccination for the covid-19 is present in [3].

In this way, if we consider the social media as an environment in which cybercrimes are developed, the echo chamber effect will be present, making the polarization of the users an opportunity for cybercriminals promote or organize crimes. An example of taking advantage of the echo chamber phenomenon by terrorist organization to afraid the population [7] or recruitment [9] is the ISIS social media propaganda.

The amount of information and the way in which it is presented and distributed makes essentially to use technological tools that facilitate the examination of the different publications that occur on the social networks. Due to, these data come from language generated by users of social media, the most convenient idea is to apply techniques that allow us to analyze, categorize and organize what is extracted from these sources of information automatically. For this, the most appropriate set of tools for this task are Natural Language Processing.

3. Objectives

The principal objective of this degree project is study and apply state-of-the-art Natural Language Processing (NLP) techniques in a counter-terrorism scenario, leading to the development a confident and scalable system of detection and prevention of terrorism related crimes. All the development is aim to the use by a Law Enforcement Agency (LEA). More specifically, the main goal may be divided into the followings sub-objectives:

1. Analyze the current state-of-the-art NLP techniques in order to choose the most suitable to the choosen application scenario.
2. Identify the principal works related to the detection of cyberterrorism.
3. Propose a scalable architecture for the proposal system.
4. Study the feasibility of the standard proposal in different real-case scenarios.

4. Methodology

Taking into consideration the social networks as open source and the type of data that would be analyzed, the set of techniques that were chosen come from the NLP. An gradual study of the techniques in NLP state-of-the-art was essential to get knowledge in order to determine the more useful tools to prevent cyberterrorism and extract actionable data. The implementation of initial models with example datasets give and excellent approach to the scope of each model. Finally, a heuristic knowledge of all the models used in NLP tasks give the possible to choose the more suitable sets of model and a idea to integrate them in order to face a cyberterrorism scenario.

With the idea of generate a NLP solution that help in the prevention of cyberterrorism a solution architecture will guide the construction of the models. The initial architecture is not going to be perfect, the best way to prove their feasibility and reliability is making test with real data that the architecture will face in production, get improvement at each test and iterate making the opportune changes. Consider all of these variables the following methodology was apply to achieve the final product of this degree project.

4.1. First Semester

a.1 Review different models of Natural Language Processing that can be applied to the detection and prevention of cybercrimes, particularly cyberterrorism.

a.2 Understand and replicate the implementation of the semantic similarity model of the proposal of «*Detecting cybercrimes using similarity models*» [10].

a.3 Identify improvement opportunities to model of [10] in their components and expand their scope. These improvements will result in a proposed architecture.

a.4 Carry out an implementation and experiments for the proposed architecture in order to obtain a minimum viable product (MVP).

4.2. Second Semester

a.5 Carry out extended experiments with real data to detect possible acts of cyberterrorism.

a.6 Apply settings to the initial proposal in order to get improvements derived by the experimentation.

a.7 Create a repository of the project that contains the implementation and their documentation to ensure the replicability of the proposal.

a.8 Write and research article that expose the achievements of the proposal and the results obtained.

This methodology was opportune in the development of the degree project because: i) Take in consideration the state-of-the-art NLP tools for the construction of a solution architecture. ii) The integration of the NLP models are based in the knowledge of the models capabilities and scope, after an example implementation. iii) The solution architecture was under constant test and improvement to achieve the a reliable solution. iv) This methodology include two ways to ensure the replicability of the project, the repository with documentation and the research article that summarize the results.

5. Results

5.1. First Semester

5.1.1. Natural Language Processing review

This activity is related to a.1 in the methodology. This study was done by completing the Natural Language Processing course offered by the National Research University Higher School of Economics, where the following applications of the subject were reviewed:

- **Sentiment Analysis:** It is used in Natural Language Processing (NLP) to extract the character of the sentiment or an opinion from the transmitter or redraw the text.

- **Name Entity Recognized NER:** In this task, a NER model is in charge of labeling each of the tokens. of a text, where labels are entities such as people, places, songs, organizations, etc.

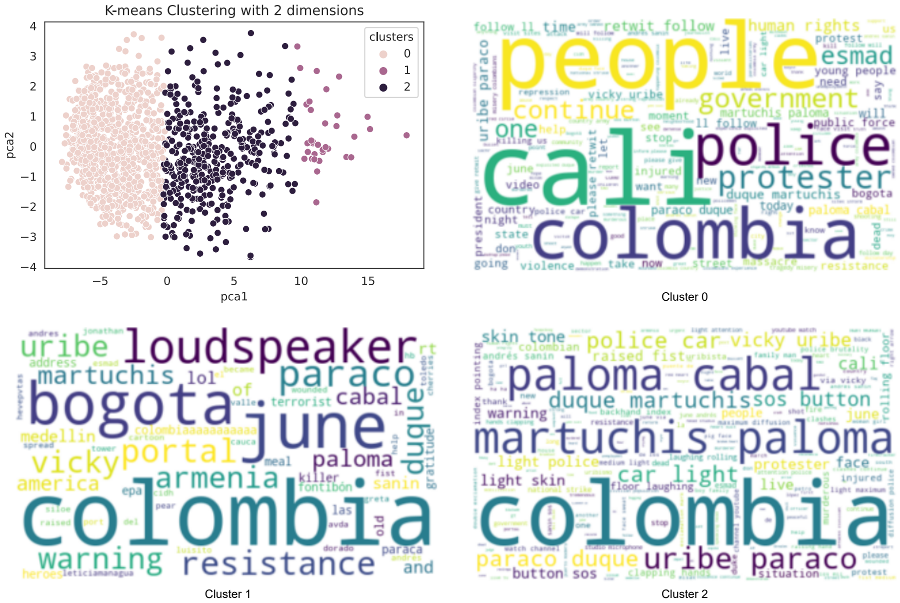
- **Word Embedding and Topics Models:** Word Embeddings is a vector representation of words, sentences or documents in the case of similarity models, this representation reproduces the meaning of the words in a real vector space. This representation is used by Topic Models to search for documents that talk about similar topics.

- **Machine Translation:** In this task we receive a source text in a particular language and it must be translated into a target language, this can be using different techniques, such as a rule-based approach, a direct translation of each word, etc. The best-performing technique is to use a neural network to generate a sequence of words in the target language using the source text.

- **Conversational models:** Using Natural Language Understanding (NLU) this task can be divided into two different subtasks, the first task is to understand the intention of the human interlocutor, the second is to generate an appropriate response to continue the conversation.

5.1.2. Profiling Twitter User Accounts

This activity correspond to a.2 of the methodology. This profiling was carried out with the capture of tweets related to the protests in Colombia against the 2021 tax reform. The analysis of these data using a similarity model allowed the grouping of individuals who had similar opinions on the subject of the protest. The groups obtained after representing the tweets in a vector way from embeddings are those shown in the Figure 1



Cluster 0

Cluster 1

Cluster 2

Figura 1: Clusters of Twitter accounts

This was followed by an analysis of sentiments to obtain the users with the greatest negative tendency, which made it possible to profile possible users who promoted violence on Twitter. After obtaining these users, a network of followers was generated and central nodes were extracted in the network that can be of interest, this can be seen in the Figure 2.

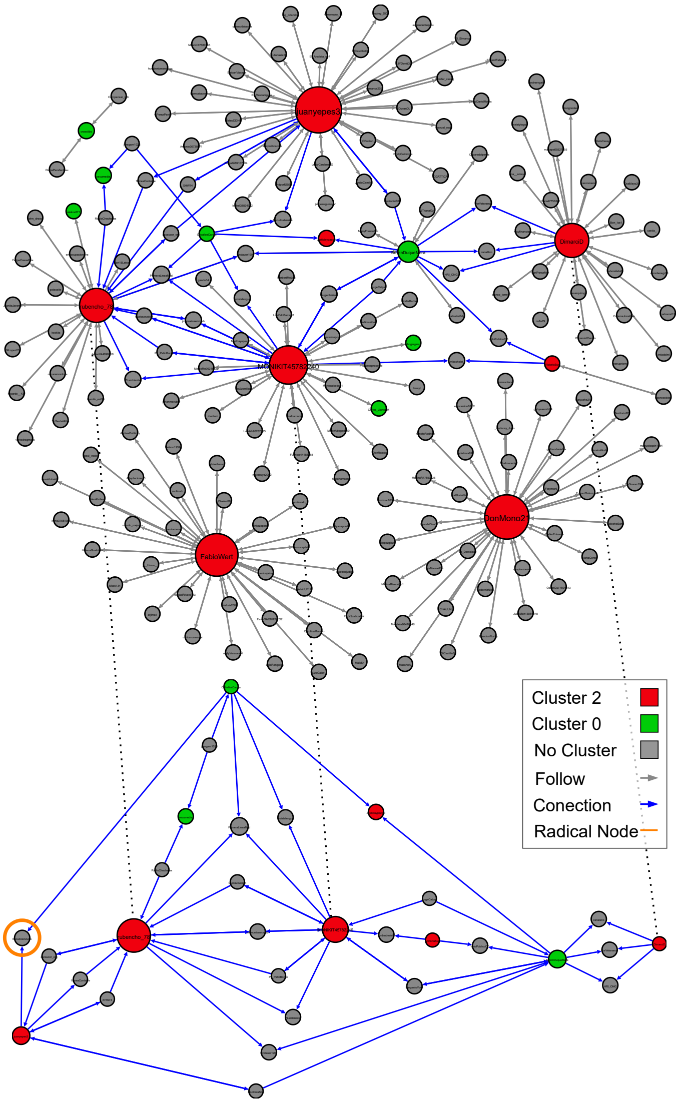


Figura 2: Graph of nodes belonging to cluster 2 showing commons followers

5.1.3. Initial Architecture

This activity correspond to a.3 of the methodology.

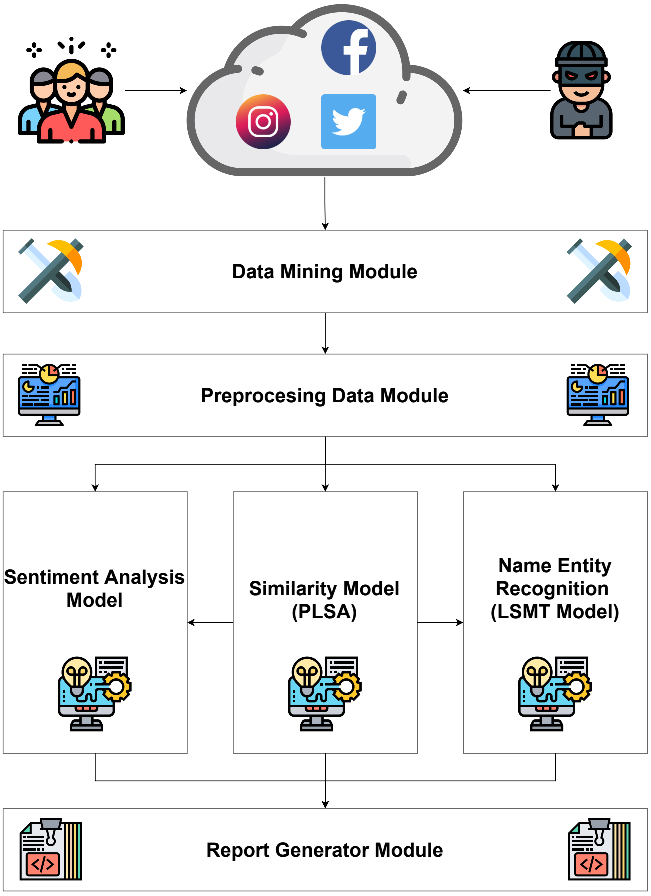


Figura 3: Initial Architecture

This architecture was designed to extract data from open sources, mainly from social networks, this information will be processed by three different components, the first of which is the similarity model mentioned in section 5.1.2, which will generate a Threat profile for each user who has made a publication in the studied social network, followed by this, a sentiment analysis model is applied that generates an alert when finding a publication that promotes violence, finally, the NER model will extract the entities found in the text to identify possible targets or sources of violence in the publication.

5.1.4. Initial architecture test

This activity correspond to a.4 of the methodology.

Radicalization detection: For this, a classifier was implemented that sought to differentiate hate posts, this classifier was fed with a database from a social supremacist forum, the implemented model achieved an accuracy of around 70 % detecting hate posts.

Entities of Interest Recognition: For this test, a NER model was implemented, training with a set of data from the darkweb focused on the sale of weapons and drugs, the results of the model were not very promising, despite this, the need for a larger database to be used was recognized, the second test that was carried out with this model was done using a tagged database of general domain tweets, the results in this case had a large improvement with respect to the first test, obtaining good results for categories as a company and geolocation, even so the precision results are not very high in other categories such as product or TV show.

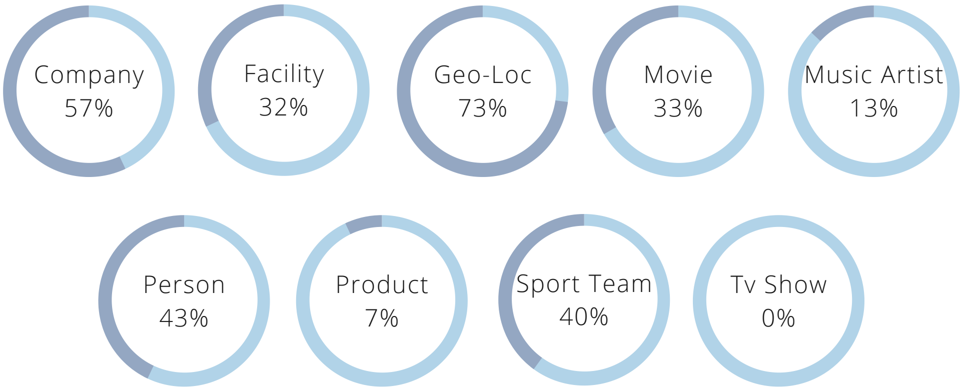


Figura 4: Results NER model tagging tweets

5.2. Second Semester

5.2.1. Extended experiments

This activity corresponds to a.5 of the methodology. The experiments were developed to test the initial architecture. In order to generate improvements to this architecture were developed using a set of tweets collected related to the protest in Ecuador on the 26th of October 2021. The three models proposed were tested with the tweets, first an adequate preprocessing was made to transform the tweets on the input format for each model. After that, the tweets were represented as vectors using two Word2Vec embeddings, an English one and a Spanish one, for the similarity model. The tweets were grouped using a similarity matrix that was built with the distance between the vector representation of the tweets. Then, three different sentiment analysis models were used to extract the polarity of the tweets. An analysis of the polarity in each cluster was made to find the most negative cluster. Using the most negative cluster a sample of users that publish the most negative tweets were chosen to get a network of contacts that allow getting analysis of the communities in Twitter. Finally, the NER model was used to extract the entities from the tweets and a review of the most common entities of each type was made. To see the entire results go check the annexes.

5.3. Final Architecture

This activity correspond to a.6 of the methodology. The final architecture in Figure 5 is composed by the same models and include a graph database that will store the data collected with the respective results of the analysis of each model and the relations between the results. Also, the sentiment model was disengaged from the similarity model to get information about each model in a independent way and increase the modularity of the architecture, this allow remove, add or change models to the solution without affect the entire functionality.

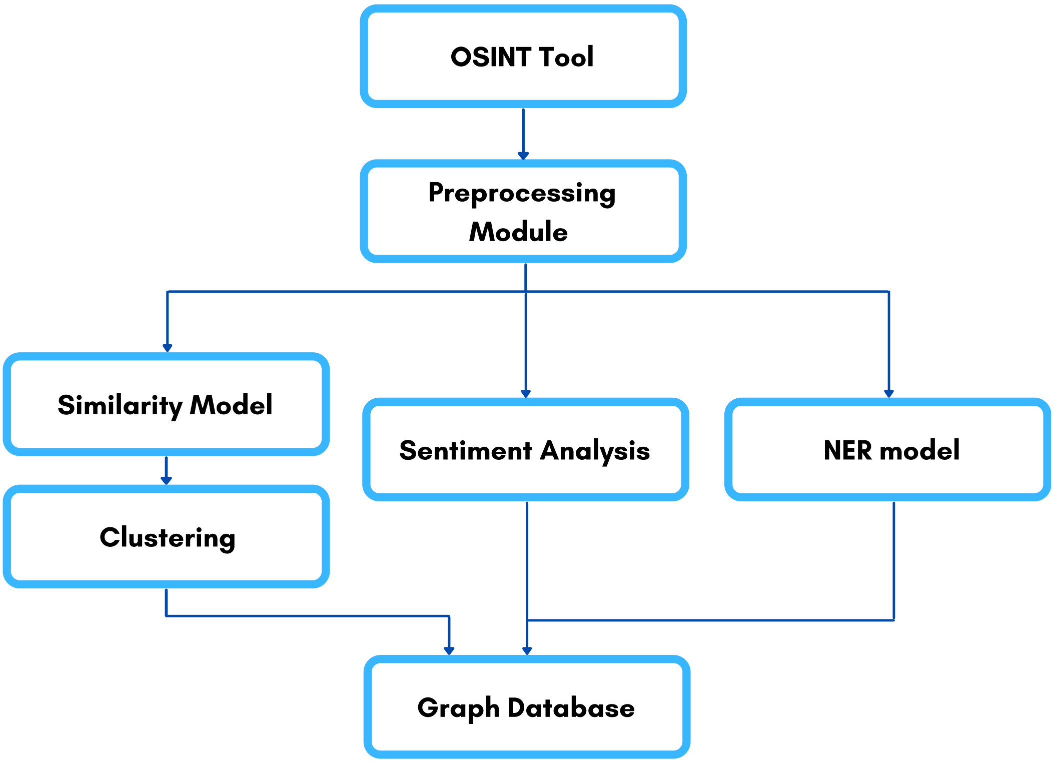


Figura 5: Final Architecture

5.3.1. Documentation and presentation of the results

This activity correspond to a.7 and a.8 of the methodology. A GitHub repository was created with the structure of the project. All the models are included as well as their training data used and the collected tweets used to test the model. This repository are available at `https://github.com/AndZapCod/NLP_Cybersecurity_Case`. Finally, the entire detail of the implementation and the results are included in the research article in the Annex section.

6. Conclusion and Future Work

- NLP models may support labors of monitoring and detecting suspicious activities expressed through human language, like hate promotion, violent speech or systematic terrorism.

- A Name Entity Recognition model may be useful to detect special entities (organization, places, persons, among others) that are being named as part of a cyberterrorism activity.

- A sentiment model may be useful to detect objetivity and emotions included in public posts or messages that allow to anticipate a hostile behavior.
- A similarity model may be useful to detect groups of individuals with similar speachs and patterns that suggest the promotion of criminal activities using social networks.
- Combination of NLP models in a cyberintelligence architecture that is algorithm-independent and extensible would allow to join forces in the fight against cybercrimes.

Referencias

[1] Akhilesh Chandra and Melissa J. Snowe. A taxonomy of cybercrime: Theory and design. *International Journal of Accounting Information Systems*, 38:100467, 2020. 2019 UW CISA Symposium.

[2] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), 2021.

[3] Alessandro Cossard, Gianmarco De Francisci Morales, Kyriaki Kalimeri, Yelena Mejova, Daniela Paolotti, and Michele Starnini. Falling into the echo chamber: The italian vaccination debate on twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):130–140, May 2020.

[4] Glenn Alexander Crowther. The cyber domain. *The Cyber Defense Review*, 2(3):63–78, 2017.

[5] Marcello Di Filippo. *The definition(s) of terrorism in international law*. Edward Elgar Publishing, Cheltenham, UK, 2020.

[6] Connor Huff and Joshua D. Kertzer. How the public defines terrorism. *American Journal of Political Science*, 62(1):55–71, 2018.

[7] Claire Seungeun Lee, Kyung-Shick Choi, Ryan Shandler, and Chris Kayser. Mapping global cyberterror networks: An empirical study of al-qaeda and isis cyberterrorism events. *Journal of Contemporary Criminal Justice*, 37(3):333–355, 2021.

[8] Marco Marsili. The war on cyberterrorism. *Democracy and Security*, 15(2):172–199, 2019.

[9] David McElreath, Daniel Doss, Leisa McElreath, Ashley Lindsley, Glenna Lusk, Joseph Skinner, and Ashley Wellman. *The Communicating and Marketing of Radicalism: A Case Study of ISIS and Cyber Recruitment*, pages 631–653. 01 2020.

[10] Andrés Zapata, Daniel Díaz-Lopez, Javier Pastor-Galindo, Félix Gómez, Julián Ramirez, and Alejandra Campo-Archbold. Decysmo: Uncovering cybercrimes in social media through similarity models. *Under review*, 2021.

Annex

- Andrés Zapata Rozo, Daniel Díaz-López. FCTNLP: Fighting cyberterrorism with natural language processing, 2021.

FCTNLP: Fighting cyberterrorism with natural language processing

Andrés Zapata Rozo, Daniel Díaz-López Universidad del Rosario
School of Engineering, Science and Technology
Carrera 6 # 12 C - 16, Bogota, Colombia
andresf.zapata@urosario.edu.co, danielo.diaz@urosario.edu.co

In the last times, the preoccupation about cyberterrorism is ground up, for instance in the U.S. is at the top of a list of 11 potential threats followed by the development of nuclear weapons by other countries and international terrorism[1]. In a cyberwar context, Hostile Social Manipulation (HSM) is a cyberterrorism strategy that employs different manipulation methods mostly through social media to produce damage to a target state. The efforts to fight cyberterrorism could be accompanied by new technologies that allow a faster and more effective contention of offensive actions. For that reason, this paper proposes an artificial intelligence-based solution that processes posts in social networks using Natural Language Processing (NLP) techniques, particularly three models: i) Sentiment Model to discriminate between threat and non-threat publications, ii) Similarity Model to identify suspects with similar intentions, and iii) NER model that identify entities in the text. Finally, the proposal was tested exhaustively to validate its functionality and feasibility, achieving an integrated and simple prototype.

cyberterrorism, Natural Language Processing, OSINT, Semantic Similarity, NER, Sentiment Analysis.

I. INTRODUCTION

Cyber-space offenses are either committed against the integrity, availability, and confidentiality of computer systems and telecommunication networks, or they consist of the use of such networks of their services to commit traditional offenses, this is also known as cybercrimes [1]. Following the previous idea we can consider terrorism a traditional crime that also migrates to a cyber-space context, this cybercrime is called cyberterrorism. Terrorism includes criminal acts intended or calculated to provoke a state of terror in the general public, a group of persons, or particular persons for political, philosophical, ideological, racial, ethnic, religious or other nature purposes[2].

When this kind of behavior involves social networks, telecommunication systems and another cyber context, then it can be considered cyberterrorism and have the same consequences as traditional terrorism. [2]. One of the best-known cases of cyberterrorism began when ISIS posted a video on YouTube on August 19th 2014 entitled "A Message to America" in which the journalist James Foley was beheaded as a response to the authorization of offensive actions against this terrorist group by the Obama's government[3].

Another case of intimidation occurred on February 10th, 2020 the Colombian guerrilla ELN, through the squadron "Omar Gomez", announced an armed strike on the main roads of Colombia to be realized in the middle of that month[4]. Those announcements were accompanied by publications on social networks. This armed strike intimidated the population and forced them to stay in their homes, people who violated the restrictions could be victims of violence from this armed group. As consequence, many towns and cities where ELN was present stopped most of their economic activities.

Regarding to Open Source Intelligence (OSINT), can be said that it is the knowledge that can be reached using public available data [3]. The internet and especially social media have contributed to the growing importance of public information that can be extracted using OSINT tools. Also, these intelligence sources have been relevant for the defense enterprise due to their potential use in big data [4] [5].

The OSINT can be complemented using Natural language processing (NLP) that's compounds computer science and linguistic to generate an approach to the understanding of the human language by a computer, this task is carried out through tools of artificial intelligence, statistics and grammar [6]. One of the most common examples of the application of NLP is a conversational agent that uses these techniques in order to understand the interlocutor language and emulates a functional conversation, taking into consideration variables as the entities and the intention of the input text [7].

In this paper we present an OSINT solution that extracts information from social networks and other resources, then such information is processed using NLP techniques including three models: i) a similarity model that relates text with similar semantic meaning, ii) a sentiment analysis model that estimates the polarity of a sentence, and iii) a Name Entity Recognition (NER) model that recognized relevant entities in the text and the type of this entities.

All the results are integrated into a simple module that resumes the output of this model generating a report that

[1] https://news.gallup.com/poll/339974/cyberterrorism-tops-list-potential-threats.aspx

[2] https://digitallibrary.un.org/record/631639?ln=es

[3] https://edition.cnn.com/2014/08/19/world/meast/isis-james-foley/index.html

[4] https://thecitypaperbogota.com/news/eln-announces-72-hour-armed-strike-warns-of-consequences-to-travelers/23849

contributes actionable information to Law Enforcement Agencies.

Thus, the main contributions presented in this paper are are described next:

- The proposal of an arquitecture whose high-level diagram is depicted in Figure 1, that combines the OSINT capabilities for extracction of open source data with NLP techniques to obtain actionable data.
- The development of different components that compose the proposed architecture through the integration of different NLP techniques and models.
- The evaluation of the proposal through exhaustive experiments, which in turn demonstrated the feasibility of the solution in a real context.

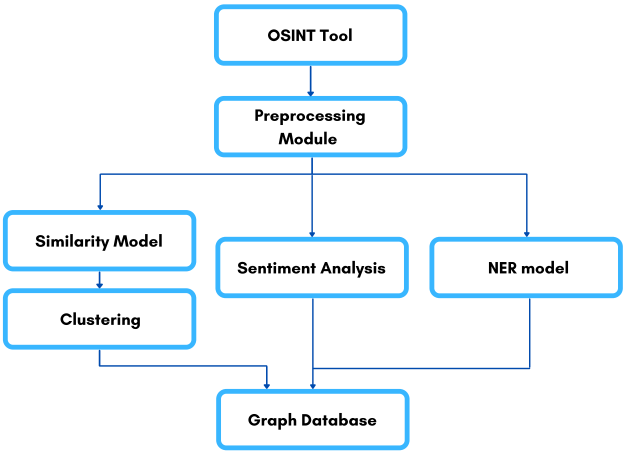


Figure 1: Abstract view of FCTNLP architecture

The remainder of this paper is structured as follows: Section II compares the major works proposed in the field, highlighting their contributions and core technique. In proposal section III, the full architecture is presented, together with its functional requirements and description of each NLP model used with the respective integration. Then, in Section IV we demonstrate the feasibility of employing the proposed architecture through exhaustive experiments. Finally, Section V concludes the work, summarizing the main outcomes and highlighting future research directions.

II. STATE OF THE ART

A compilation of works that use NLP to fight against cybercrimes is presented next. In [8] the authors use data from two datasets: one available online and another built with data from Twitter and Facebook and labeled manually to construct a cybercrime text classifier. In addition, they compare their different classifiers with sentiment analysis from the NLTK Python library.

Work proposed at [9] consists of a big data architecture that allows a real-time analysis of tweets to classify users and their respective followers as part of ISIS a terrorist organization, according to parameters like: level of activity, influence in other users, and post content. A graph was created where indicators of centrality were applied to identify the most influential users before applying analysis over the data. Finally, user profiles were obtained through Fuzzy clustering techniques.

Hate speech can be considered a type of terrorism that intimidated a specific target population. In [10] the detection of hate speech in the Arabic context was developed through the use of different comparative machine learning methods like Support Vector Machine (SVM), Naive Bayes (NB), Decision tree (DT) and Random Forest (RF). The data used in this work come from tweets related to racism, journalism, sports orientation, terrorism and Islam.

The detection of cyberterrorism vocabulary in web pages was proposed in [11] and in this work the results of the following algorithms were evaluated: Random Forest, Boosting, SVM, Neural Network, K-Nearest Neighbor (KNN) and Naive Bayes. Where a Random Forest approach gives the best results. In all the cases the percentage of accuracy was higher than 80% and in the case of the Random Forest approach was 95.62%. The vocabulary developed to detect the websites include information related to Al Qaeda, Supreme Truth, KKK and ETA groups.

An analysis of a historical dataset was present in [12] where the data from Twitter messages of attacks between 2008 and 2019 of the terrorist group Boko Haram in Nigeria were analyzed using DynamicK-reference clustering algorithms. This work allowed identify various of the strategies of Boko Haram attacks and pointed out weaknesses in security control in sectors of northern Nigeria.

The proposal presented at [13] tries to detect cyberterror and extremism in the text using Fuzzy sets-based weighting methods, Naive Bayes Multinomial (NBM) and SVM. The experimental analysis shows that the fuzzy set-based weighting method with SVM classifier gives the best classification with a 99.4% of accuracy.

Table I shows a comparison between the related works described in the previous paragraphs taking into account aspects of the proposal goal, the core technique employed, the used dataset, its dimension and language. As seen in Table I the most recent works that make contributions in the fight against cyberterrorism through the use of NLP has the purpose of the detection of terrorist behavior on the internet.

Proposal	Purpose	Core Technique	Dataset used in experiments	Dataset Composition	Dataset Language
[8]	Compare different models to differentiate cybercrime in Twitter and Facebook content using sentiment analysis	NB, MNB, Bernoulli NB, Logistic Regression, Stocastic Gradient Decent Classifier, Linear Support Vector Classification (SVG), Nu-SVG to classify between 'positive' and 'negative' content	Own	3000 Tweets	English
[9]	Classify users on Twitter according to their activity, influence and post content	Fuzzy Clustering that allow different users to belong to different groups simultaneously, but with different membership degree	How ISIS Uses Twitter	17410 tweets	English
[11]	Detection of cyberterrorism in web pages	Random Forest model to classify web pages that contain vocabulary related to cyberterrorism	Own	NA	Spanish
[12]	Analyze and historical dataset of the Boko Haram insurgency attacks between 2008 and 2019	DynamicK-reference algorithm to group the different reports using the categorical features of the dataset.	Own	3000050 Tweets and Facebook reports	English
[13]	Detection of Cyber terror and Extremism in text	Hidden Markov Model and SVM over Binary, TF, TF-IDF and Fuzzy sets based weighting to represent the text and classify the text as cyber terror/Extremist or not.	Antisocial Behavior dataset (ASB)	148 Documents with 680 characters in average each one	English
[10]	Hate Speech detection in Twitter	SVM, NB, DT, RF to classify tweets as hate speech or not.	Own	3696 tweets (843 hate, 790 non-hate, 2061 neutral)	Arabic
Our Proposal	Cybercrimes prevention on Twitter	Similarity model to get related topic in tweets, K-Means clustering and Sentiment Analysis to group users and detect suspicious post, NER model to identify, locations, organization, persons, etc. in the text and Graph database to link the data and analysis.	Wikiner to train NER model, Own to get the results	Wikiner-es: 7,200 Wikipedia articles Tweets recopiled: ??	Spanish

Table I: Related works

III. FCTNLP

This section describes the main aspects of the design of FCTNLP, covering the definition of requirements and the explanation of the main components. The development of this design follows the phases defined in a data science life cycle [14]: i) Business understanding, ii) Data acquisition and preprocessing, iii) Modeling, and iv) Deployment.

A. Business understanding

As seen in Section I cyberterrorism is a real problematic that affect the general population and as a crime it should be fought. Following this, One of the main problems in the fight against cyberterrorism is recognizing it in the middle of a large data flow, such as the one existing in social networks. In addition to the amount of data that must be analyzed, human-generated text may display different structures with different meanings [15]. Thus, one option to analyze sentences is to devote a person to extract key information from such human-generated text, however that analysis would be subjective and a very large group of people would be needed to analyze all content coming from social networks. This is why a solution like FCTNLP that automates the structuring and analysis of posts on social networks is vital in the fight against cyberterrorism. Thus, the architecture proposed of FCTNLP is expected to meet the following targets:

- Distinguish tweets: It should be capable of conform initial groups of tweets related by their semantic meaning.
- Evaluate polarity: It should contain a model that score the polarity of each tweet.
- Recognize entities: It should be capable of extract entities relevant to track cyberterrorism.
- Identify communities: It should use OSINT information related to the Twitter accounts that are generating content to identify existing relations between such actors.
- Extensible: It should allow the integration of more data science models in the future to provide better and more extensive information.
- Model Independent: Regardless of the implementation of the models, the architecture should offer the same functionalities described previously.
- User friendly: The end user (cyber intelligence agent) of a solution that implement this proposed architecture must be able to interact with even if he does not have a substantial knowledge of NLP.

B. Data acquisition and preprocessing

Amongst all social networks, Twitter has consolidated as an important source of data due to the relevance and diversity of data that can be obtained from it to be analyzed [16]. In terms of the impact of tweets in the population perceptions or actions it is interesting to see how the tweets of Donald Trump impact in the American Democracy [17] or how Elon Musk's tweets moves the crypto currency markets [18]. Thus, we consider Twitter as a social network with the possibility to generate a high impact beyond the cyberspace that is why such social network was selected in this paper as the source of the raw data that feed our proposal.

Different OSINT tools and techniques may be used to gather tweets, which can be divide in two categories: i) Scrappers and ii) API based tools. The first category contains scrappers that uses bots, some of them emulating human behavior, to get specific data from Twitter, e.g. Octoparse[5]. This kind of tools allow to personalize the type of data that will be extracted but for policies of Twitter most of these kind of tools have a tweets extraction limit of 1000 tweets per day. These tools in general can be set to get different kind of data from different web pages, not only Twitter. The second category refers to tools that consume the Twitter API[6] and therefore run under the restrictions defined by such API, e.g. tweets can only be extracted from a short window of time that may be up to the last 7 days previous to the date of the collection. Some examples of API based tools that allow to collect tweets and even to apply advanced filters to reduce the data to be studied are Mozdeh[7] and TAGSv6.1[8].

After tweets are collected, many strategies of preprocessing can be used to prepare the text before feed the NLP models. The principal objective of these strategies is keep the meaning of the text but clean it from noise data that can influence in a bad way the performance of the models. Between the most common strategies applicable to tweets are: i) remove hashtags, mentions, URL's, strange characters and punctuation symbols, ii) normalize text that convert text in lower or upper case, iii) replace emoji's for words that represent its meaning, iv) apply tokenization that divides the text in tokens that can be only words or words with punctuation symbols, and v) do lemmatization that replaces a word by their lemma, i.e the canonical form of the word. The different strategies used in the implementation will depend on the NLP model that will be fed with such preprocessed data.

C. Modeling

In this section three NLP models will be described, the first one is the sentiment analysis model that is used in the NLP tasks to determine the emotions that the author of a text expresses or the mood of the author at the moment to write the text. Secondly, a NER model that is used in NLP to extract relevant entities and their respective type from a raw text. Finally, the similarity model that is used to represent a text in a vector way so that the representation can contain the semantic meaning of the text.

The inclusion of these models allow to achieve the targets proposed in section III A in the following way: i) the similarity model allows to distinguish tweets in groups according to its semantic meaning, ii) the sentiment analysis model allows to evaluate the polarity of the tweets, iii) the NER model allows to recognize entities relevant to cyberterrorism activities.

1. Sentiment Analysis Model

A sentiment analysis model can be implemented as a classifier which discriminates text between classes according to the polarity of the text (positive, negative or neutral), that classifies the subjectivity of the author (subjective, objective), or that extracts the emotional state of the text (happy, angry, friendly, confident, etc.) [19].

Sentiment analysis models may also in a single function, e.g. regression function, scores two or more aspects of the text like the polarity and subjectivity [20]. Finally, another way to implement a sentiment analyzer is using a ruled-based algorithm using the knowledge about the language structure and the meaning of the words [21].

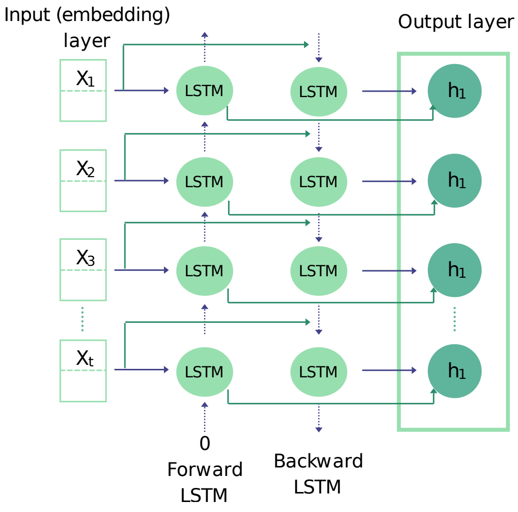


Figure 2: Architecture of a Bi-LSTM based model

[5] https://www.octoparse.com/tutorial-7/scrape-tweets-from-twitter
[6] https://developer.twitter.com/en/docs/twitter-api
[7] http://mozdeh.wlv.ac.uk/
[8] https://tags.hawksey.info/

2. Name Entity Recognition Model

A Long Short Term Memory (LSTM) is a Recurrent Neural Network (RNN) that takes advantage of consecutive and no-consecutive terms to give the best results in the task of recognition of entities from human language. It may also offer an understanding of the relation between words and their grammatical meaning. Another architecture for this task, which is shown in Figure 2, is the Bidirectional LSTM (Bi-LSTM) where two LSTM model are concatenated in a way that the Bi-LSTM can receive information from the beginning of the text up to the end, and from the end up to the beginning [22]. Independent of the RNN used for the implementation, a NER model is generally trained using a Begin Inside Outside (BIO) notation where a phrase is decomposed in beginning, inside and outside sections. Thus, a NER model adds actionable information to analyzed tweets such as the one that helps to describe where (location) and who (subject, organization) is involved in some specific actions being monitored.

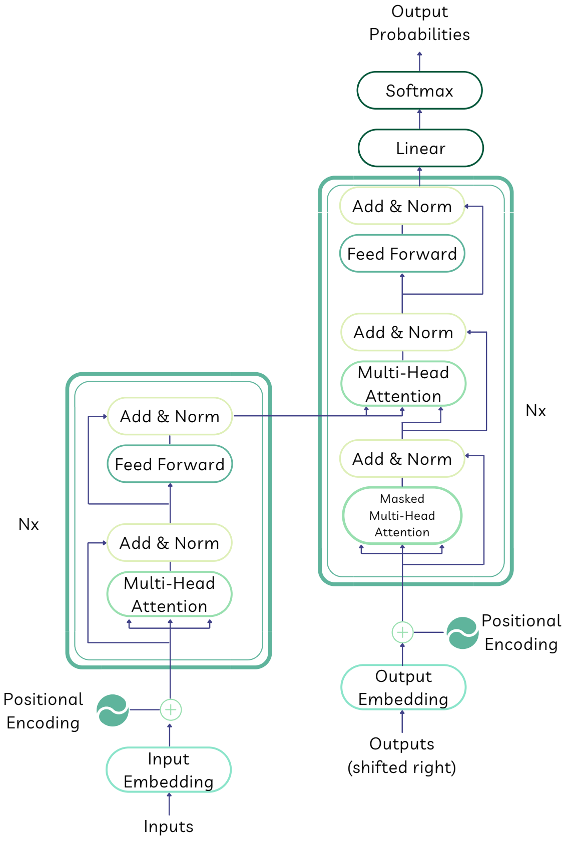


Figure 3: Architecture of transformers based model

Other architecture used in NER tasks is called transformers, as depicted in Figure 3. It was proposed in [23] and consists of two stacks: one encoder and one decoder. As shown in Figure 3, both inputs and outputs have embeddings and positional encoding. Each stack uses multi-head attention layers: a non-masked one for the encoder, and a masked one for the decoder. At the end of both stacks a fully connected feed-forward network is also placed. Finally, a linear layer and a softmax activation function are used to get the prediction.

3. Similarity Model

The similarity model uses word embeddings to represent the meaning of the words in a real space. Such representation is useful to get the relation between words that have a similar meaning as they will have a close vector representation. The metric used to calculate the similarity between words is the soft cosine distance (1) which is represented by Equation 1.

$$soft\_cosine(w,v) = \frac{\sum_{i,j}^{N} s_{ij} w_i v_j}{\sqrt{\sum_{i,j}^{N} s_{ij} w_i a_j}\sqrt{\sum_{i,j}^{N} s_{ij} v_i v_j}} \quad (1)$$

To generate the word embeddings two principal algorithms may be used: Word2Vec and FastText. In the case of Word2Vec it creates a vector representation for each word in the text, keeping similar words close in the vector space [24]. This approach has the problem that words that are not included in the training set (new words included in tweets) will not be considered in the similarity calculus as they will not have a vector representation. FastText algorithm may help to solve this previous problem as it uses a vector representation that takes in consideration the n-grams of a word, i.e. the sequence of n characters. Thus, this last approach is capable of represent words that are composed by some n-grams contained in the training dataset, even if it losses the property of have semantic knowledge in the vector representation [24].

An illustration of how the previously described models (similarity model, sentiment analysis model and NER model) are integrated in the FCTNLP architecture is shown in Figure 4.

D. Deployment

An implementation of the FCTNLP architecture could be used by a Law enforcement Agency (LEA) to automatize the analysis of a human text obtained from open sources and help in the prevention of cyberterrorism. Such implementation can be a key tool for an cyber intelligence agent as it facilitates the search of spotlights of cyberterrorism. Results obtained from FCTNLP may also be enriched with information provided by an already running commercial cyberintelligence solution.

Due the modularity, the high cohesion and the low coupling of the proposed architecture, each NLP model can be deployed as part of a scalable solution that allows

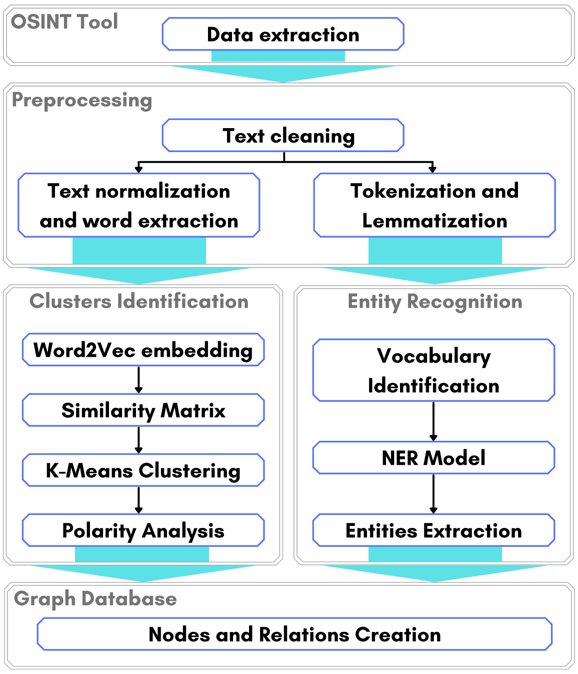


Figure 4: Components of the FCTNLP architecture

the flow of a huge amount of data, for instance to be deployed as a micro service. The extraction of tweets can also be set as a task to be executed on real time or under demand. In both cases, extracted data can be saved in a data lake that will be processed by the NLP models that compose FCTNLP, and a cache solution can also be fed with the more recent and relevant results in order to have a quick access to required data.

IV. EXPERIMENTS

This section contains the results obtained from applying the proposal described in Section III in a scenario related to a protest that occurred on October 26, 2021 in Ecuador, being the data and code available at the project repository [9]. Twitter was the social network used to provide the raw information to be processed. The gathering was done using `TAGSv6.1`[10].

The embedding process in the English language was based on the use of Google News Embedding, which contains generic embeddings for 3,000,000 English words with dimension 300. On the other hand, the embedding of content in Spanish was done using the Spanish Billion Words Corpus and Embeddings[11] that contains a set of 1,000,653 words with 300 dimension vector representation. both embeddings were built using the word2vec algorithm word2vec[12] that was described in section III. The extraction of information related to followers from Twitter accounts was made using tinfoleak[13] and the generation of the neighborhood graph for a selected cluster was made using Gephi[14].

A. Gathering tweets in the protest against economic policies in Ecuador

The situation that broke the camel's back was the announcement of the increase in gasoline prices by the government of Guillermo Lasso[15], in addition to the economic crisis in which Ecuador finds itself and the fall in popularity of the Lasso government due to the investigation that he is facing due to he is appearing involved in the Pandora Papers[16].

This scenario implied the gathering of 10,608 tweets containing at least one of the following hashtags `#ParoNacional`, `#ParoNacionalEC`, `#LassoEsUnFracasso`, `#Quito`, `#LassoCorrupto`, `#LassoMentiroso` and in order to obtain only original tweets with text we use the following filters of the twitter API `-filter:images`, `-filter:videos`, `-filter:retweets`. Between October 24 and October 27 of 2021. The protest occurred between 26 October and October 27, 2021, for this reason the tweets before these dates were omitted so a total of 7,086 tweets remained.

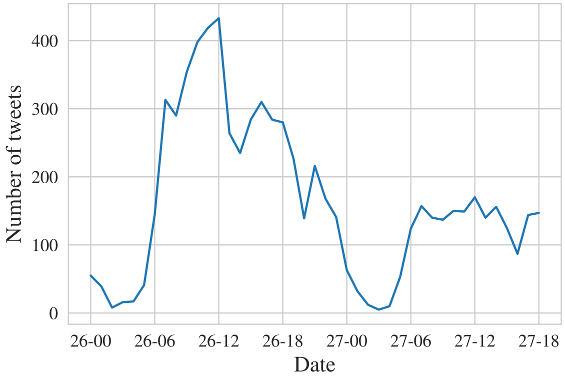


Figure 5: Number of tweets posted before, during and after the protest day

[9] https://github.com/AndZapCod/NLP_Cybersecurity_Case
[10] https://tags.hawksey.info/
[11] https://crscardellino.ar/SBWCE/
[12] https://code.google.com/archive/p/word2vec/
[13] https://tinfoleak.com/
[14] https://gephi.org/
[15] https://www.argusmedia.com/en/news/2266703-ecuador-freezes-fuel-prices-update
[16] https://www.reuters.com/world/americas/ecuador-president-lasso-be-investigated-tax-fraud-after-pandora-papers-leak-2021-10-21/

In the Figure 5 we can see the amount of published tweets between October 26 and 27. In this graph can be appreciated the time interval in which the protests took place it is where the graph contains its higher values.

The day of protests was marked by some acts of violence, the most serious of which was the confrontation between the demonstrators and the police in front of the presidential palace[17]. On the other hand, other disturbances occurred in various parts of Ecuador such as road blockades[18]. At the end of the protests, 37 people were arrested for acts of violence[19].

The data of this experiment is a unique collection from Twitter that uses TAGS. As we see in the section III tools like TAGS that used the Twitter API to collect the information from Twitter have a limit to the tweets that can be collected in a window of time. Most of the time this limit is not reached, especially with specific topic queries like the use in this experiment.

Tweets were preprocessed and cleaned properly to be consumed by the models that will be used later in the pipeline. The first step in preprocessing was to construct a dictionary with the most used hashtags and mentions of the collected tweets and replace them with their meaning words. The second step was to remove URLs, mentions, hashtags, reserve words like RT and FAV, smilies and strange characters were removed from the tweets using the python library `tweet-preprocessor`[20]. Then, emoticons were replaced by their meaning in words through the use of the Python library `emoji`[21]. Finally, empty and duplicated tweets were removed and a total of 7,077 tweets remained.

Additional preprocessing was required for each model, as explained next:

- **Sentiment analysis model:** Punctuation symbols were removed from the text and the text was normalized to lowercase.
- **Similarity Model:** Tweets were translated from Spanish to English using Google API Services[22]. Punctuation symbols were also removed and the text was converted to lowercase. The purpose of this translation was to uniform the language to the one used by Google News Embedding, to be able to vectorize the Tweets.
- **NER model:** Cleaned tweets were tokenized and each token was lemmatized using a python dictionary constructed in the training of the model to translate a token to an integer to be used as input for the embedding layer in the Bi-LSTM model see at section III.

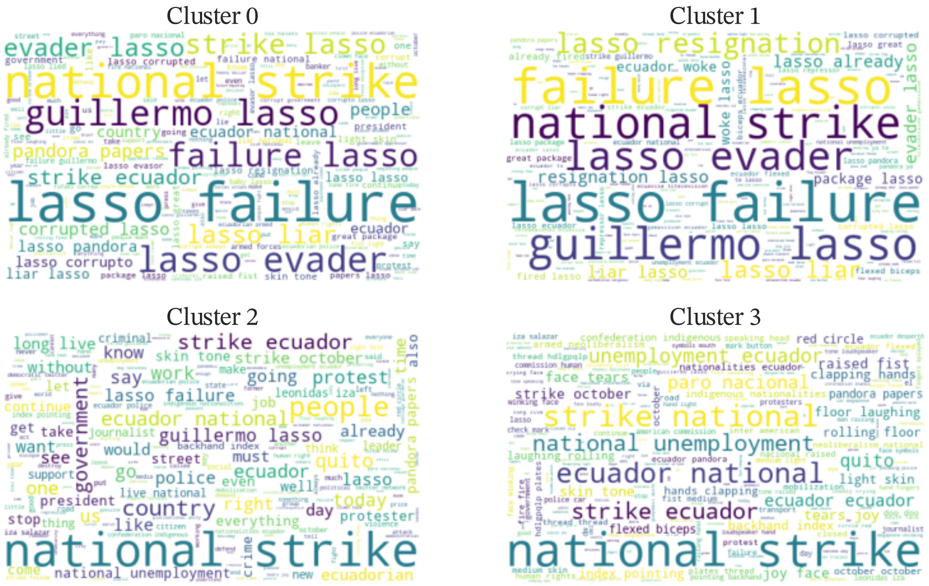


Figure 6: Word cloud for the clusters with google news embeddings

[17] https://www.laprensalatina.com/quito-violence-marks-day-of-protests-against-ecuador-president/

[18] https://frontline.thehindu.com/dispatches/protesters-in-ecuador-block-roads-over-gasoline-price-hikes/article37195681.ece

[19] https://www.reuters.com/world/ecuador-demonstrators-block-some-roads-protests-over-gas-prices-2021-10-26/

[20] https://pypi.org/project/tweet-preprocessor/

[21] https://pypi.org/project/emoji/

[22] https://pypi.org/project/google-cloud-translate/

B. Application of the similarity model

For the analysis of the translated tweets, $(t_m = t_1, \ldots, t_n)$ were processed by the similarity model mentioned in Section III using the google news embedding to build a matrix of cosine distances between the different tweets. The entries of such a matrix were done taking each tweet (t_i) and calculating their cosine distance against the remaining tweets. Afterward, The 7,077 tweets were split into four clusters according to the K-means algorithm using the similarity matrix in order to take advantage of the semantic representation of the tweets.

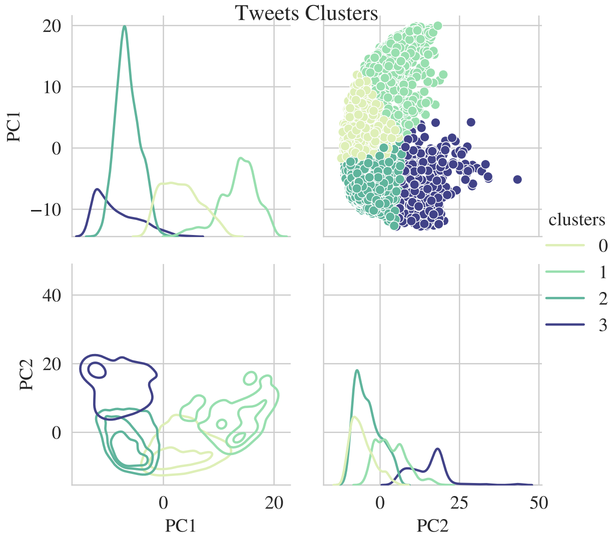


Figure 7: K-means clustering using Google News embeddings

In the Figure 7 we can see a representation of the tweets using the similarity matrix and the PCA algorithm to extract the two dimensions that represent the 86.4% of the variability of the data.

Then, the collected tweets $(t_m = t_1, \ldots, t_n)$ were processed to obtain a validation dataset composed by a tweet (t_i), their most similar tweet (t_s) and two other randomly selected tweets (t_p, t_q) that will be used as negative examples for the metrics. Thus, a dataset that contains for each row the following structure $\{t_i, t_s, t_p, t_q\}$ was composed. This validation dataset and the original dataset were compared to verify the correctness of the similarity model.

Taking into consideration the results shown in Table II, which shows that the similarity model gets better results in the Hits metric in comparison with the DCG metrics when the training interactions increase. It can be explained that the DCG metric takes into consideration the order in which similar tweets are ranked.

In the Figure 6 we can see that the reasons for the protest are dominant in all of the clusters. In addition to this, the reason for the protest, Ecuador's president is more mentioned in clusters 0 and 1.

Interactions	DCG	Hits
1	0.425	0.425
5	0.754	1.000
10	0.754	1.000
100	0.754	1.000
500	0.754	1.000
1000	0.754	1.000

Table II: Metrics DCG and Hits with Google News embeddings

For the case of the clusters that were obtained through K-Means and using the Spanish Billion Embeddings to calculate the similarity matrix, the results are contemplated below.

In the Figure 8 can be appreciated the clusters obtained for the Spanish version of the tweets, in this case, are three clusters. In the same way as before this Figure shows the principal components obtained using PCA and these components represent the $86,9\%$ of the tweets variability.

In the same way that before, the results shown in Table III show that the similarity model gets better results in the Hits metric in comparison with the DCG metrics when the training interactions increase. It can be explained that the DCG metric takes into consideration the order in which similar tweets are ranked.

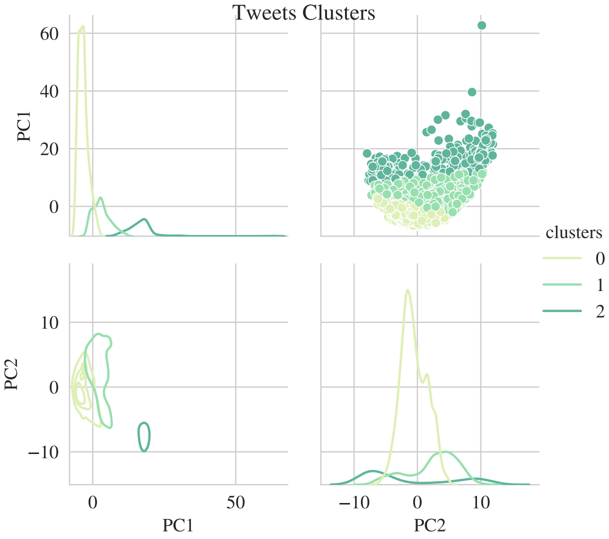


Figure 8: K-means clustering using Spanish billion words embeddings

Interactions	DCG	Hits
1	0.264	0.264
5	0.671	1.000
10	0.671	1.000
100	0.671	1.000
500	0.671	1.000
1000	0.671	1.000

Table III: Metrics DCG and Hits spanish Word2vec

In the Figure 9 we can see that the topic "national strike" is common in all of the clusters with a similar proportion of the occurrence due to it have the same size in all of the clusters. These common topics in the clusters can be explained when we analyze the Figure 8 where the vector representation of each tweet is closer than in the case of the Figure 7.

Despite having embeddings in the original language of the tweets these embeddings have close to three times fewer words at the time to represent a tweet as a vector than the google news embeddings. As we see in section III the performance of the vectors that are generated using the Word2Vec algorithm depends on the size of the vocabulary used to train the embeddings. For this reason, the English embedding can represent in a better way the tweets studied. It also can be seen comparing the DCG metric of both models.

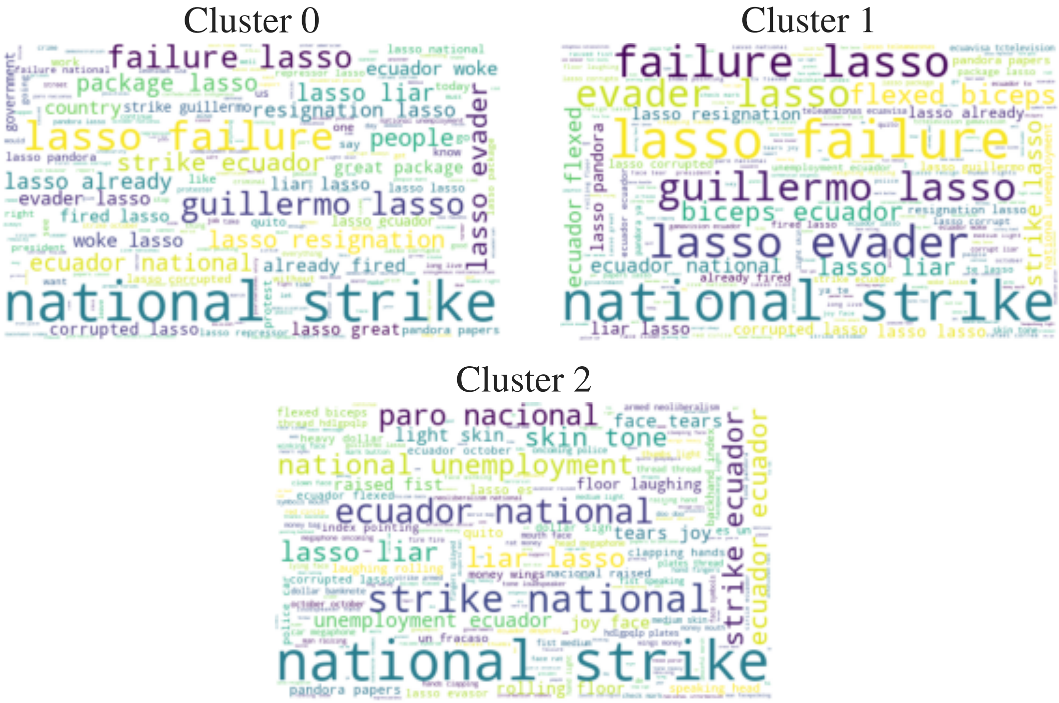


Figure 9: Word cloud for the Spanish clusters

C. Application of the sentiment analysis model

For each cluster different sentiment analysis was conducted, in the first case, the analysis was used the `TextBlob` python library. that use a single perceptron to extract a score of the polarity between $[-1, 1]$ the values with more negative score are also with a polarity more negative.

As we can see in the Figure 10 the cluster with a higher proportion of negative tweets is the cluster 1 with a 64% of negative tweets and as we can see in the Table IV the cluster 1 has also the less polarity score, i.e we can consider a cluster of tweets that may contain cyberterrorism.

For the second case, the sentiment analysis was conducted using the Vader sentiment analyzer from `NLTK` python library. That is a lexicon and rule-based sentiment analysis tool to score the polarity between $[-1, 1]$ the values with more negative scores are also with a polarity more negative.

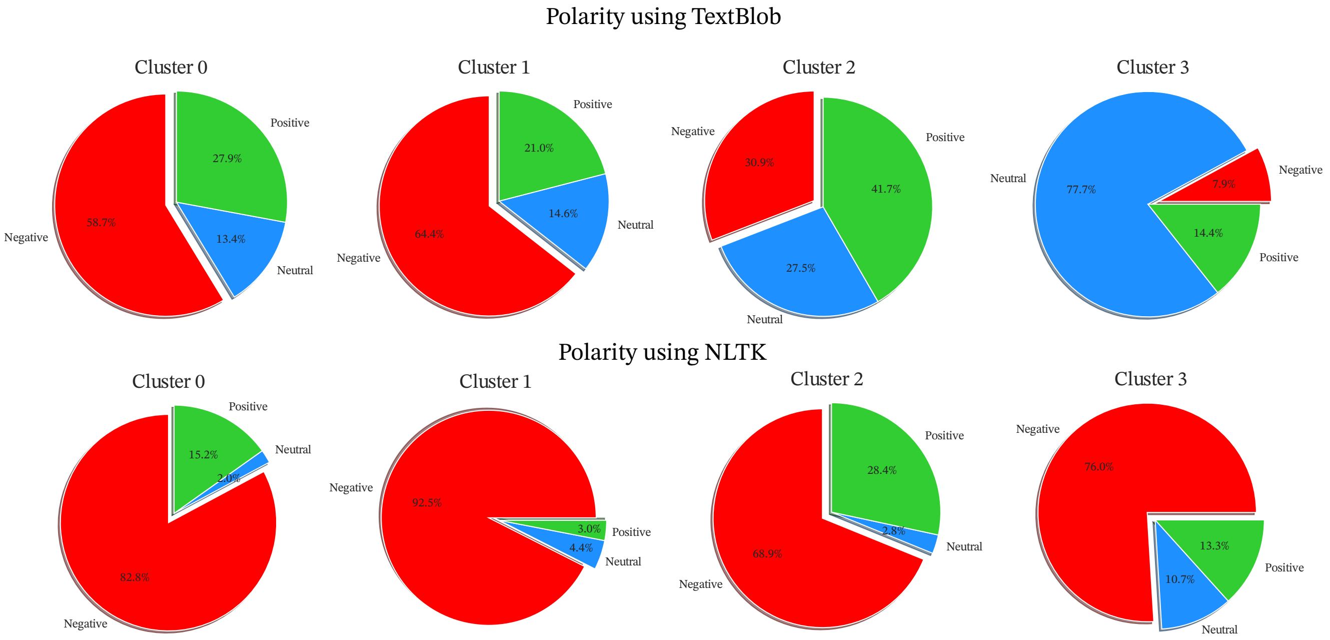


Figure 10: Polarity with google news embeddings

Similar to the results using the `TextBlob` Sentiment Analysis tool, The cluster 1 has the greater proportion between all the clusters with the 92.5% of the tweets as we can see in the Figure 10, and in the same way, the cluster with the most negative score as mean is the cluster 1 as we can see in the Table IV.

	TextBlob		**NLTK**	
Cluster	**Negative**	**Positive**	**Negative**	**Positive**
0	-0.25	0.23	-0.60	0.39
1*	**-0.33**	0.24	**-0.65**	0.39
2	-0.21	0.22	-0.51	0.42
3	-0.23	0.32	-0.26	0.50

Table IV: Sentiment analysis libraries comparison

At this point both sentiment analysis approach throws similar results at the time to recognize the more negative cluster, the biggest difference in the results of both approaches for the English sentiment analysis is the proportion of negative detection, in the case of the `NLTK` approach more number of tweets have a negative score in all of the clusters.

Finally for the clusters obtained using the billion words embeddings for Spanish another sentiment analysis model was used in this case the model was extracted from `sentiment_analysis_spanish` python library. In this case, the score of polarity is given between $[0, 1]$ in order to homogenize the score was mapped to the interval $[-1, 1]$ using a simple convex combination.

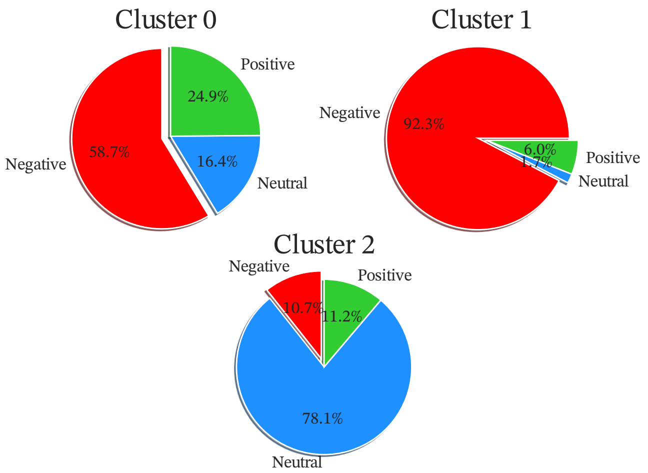


Figure 11: Polarity analysis using `sentiment_analysis_spanish`

Cluster	**Negative**	**Positive**
0	-0.65	0.26
1	**-0.90**	0.48
2	-0.63	0.38

Table V: Sentiment analysis with `sentiment_analysis_spanish`

As we can see in the Figure 11 the cluster with a higher proportion of negative tweets is the cluster 1 with a 92.3% of negative tweets and as we can see in the Table V the cluster 1 has also the less polarity score, i.e we can consider a cluster of tweets that may contain cyberterrorism.

D. Application of the NER model

In this case, a NER model was trained using the WikiNer dataset [25] with the Spanish corpus, this dataset contains 141,761 sentences extracted from Wikipedia and annotated using the BIO format with three types of entities: i) *LOC* Localization, ii) *MISC* Miscellaneous iii) *ORG* Organization and iv) *PER* Person. The model was trained using the Bi-LSTM architecture see in the section III. The metrics obtained in the train set and the test are shown in the Table VI

In Figure 12 we can see the ten entities with the highest frequency of appearance of each type of entity predicted by the NER model after remove bad predictions as stops words, punctuation symbols and names of emojis. The last name of the president of Ecuador is part of the set of entities with greater frequency in all categories of entities, in the case of entities such as person and organization, this would make sense since depending on the context this word can refer to the president as a human being or to the as a representative of an organization, in this case, the presidency. It must also be considered that this word was not found in the training set, so the model managed to label it in any case.

	Train Set Quality		
	Precision	***Recall***	***F1-Score***
General Performance	95.37%	95.98%	95.67%
LOC	94.73%	95.76%	95.24%
MISC	92.77%	93.05%	92.91%
ORG	93.88%	93.21%	93.55%
PER	98.11%	98.56%	98.34%
	Test Set Quality		
	Precision	***Recall***	***F1-Score***
General Performance	82.45%	84.93%	83.67%
LOC	82.90%	86.77%	84.79%
MISC	67.77%	68.94%	68.35%
ORG	77.48%	77.33%	77.41%
PER	89.91%	91.50%	90.70%

Table VI: Tain and Test Quality

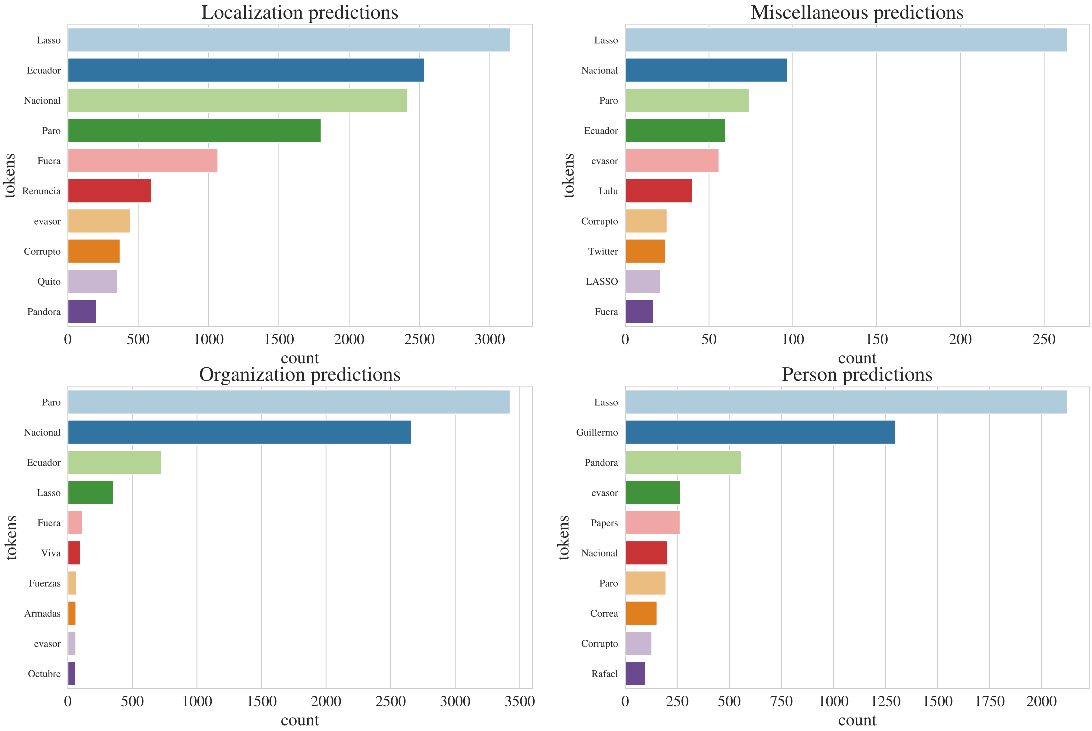


Figure 12: The ten most predicted entities for the Spanish tweets

On the other hand, correctly labeled locations were for example Ecuador, Quito. in the case of organizations: national strike and armed forces as well as its individual tokens. In the case of Person, as we can see, he correctly predicted the name and last name of the president of Ecuador, as well as the name and last name of Rafael Correa, an ex-president of Ecuador, and he labeled the leaked documents known as "pandora papers" as a person.

Although the model successfully predicts several of the entities that were relevant within the context of the protests, errors also occur that are due to the difference between the training set, Wikipedia articles, and the analyzed tweets that also contain a language more informal in addition to the use of emojis.

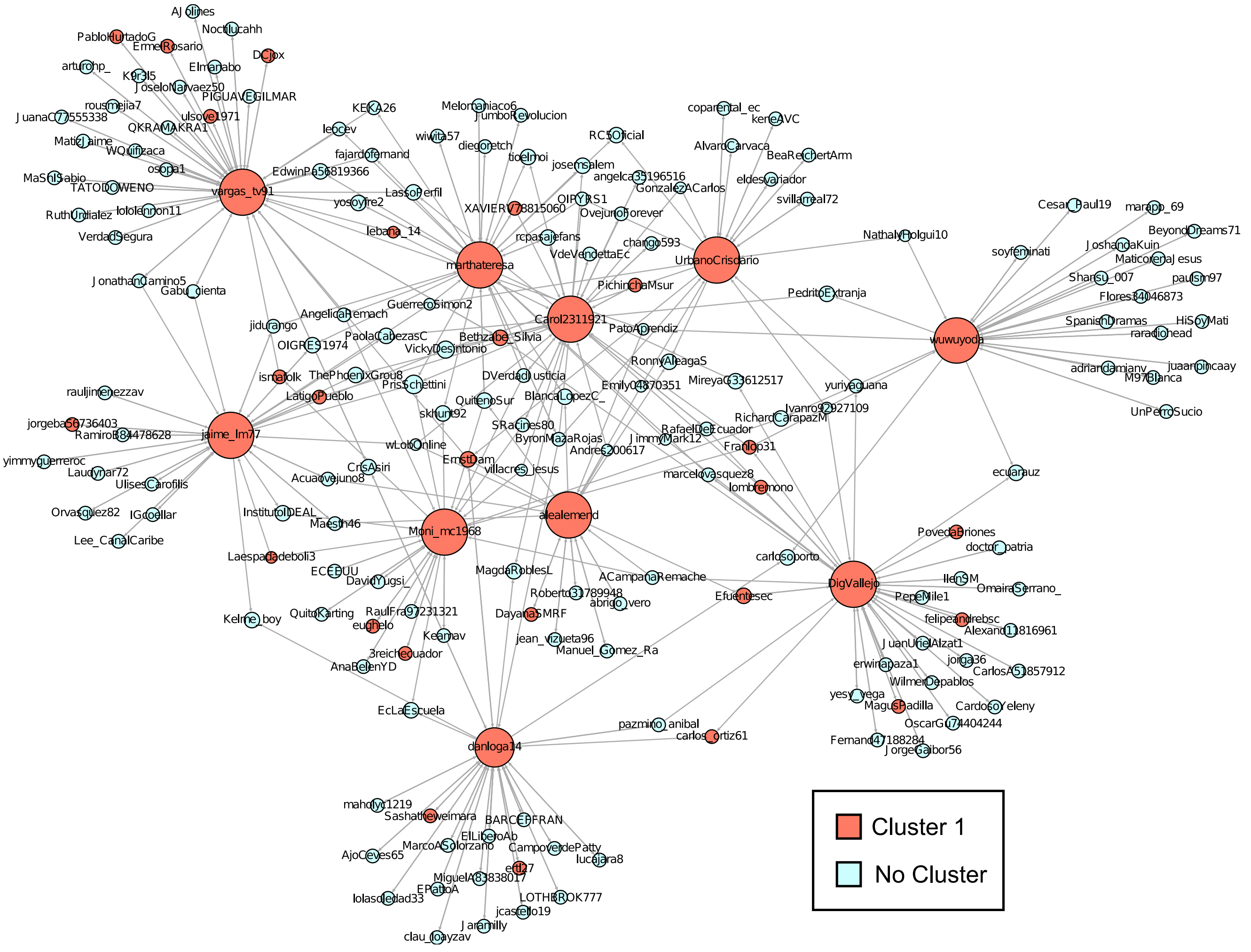


Figure 13: User network sample of user of the cluster 1

E. Graph representation of the user network of contacts

Finally, the data of the Twitter users were collected using the `tinfoleak` tool in order to construct a network of contacts In the case of finding some type of suspicious activity, and intelligence agencies can consult a graph like the one shown in Figure 14. In this case, it is a sample of users who published tweets grouped using google news embeddings in cluster 1 (red nodes), which was the one with the most negative polarity and some of its contacts (blue nodes) from which tweets not were collected.

In this graph, we can see communities of users that follow big red nodes these communities are connected by internal nodes that are common contacts between these big nodes and other red nodes.

V. CONCLUSIONS AND FUTURE WORK

Taking in consideration the influence that have the social networks in the people and the use of this media to promote or organize cyberterrorism, the use of NLP as a support on the cybersecurity prevention. It has been considered as a way to automate the analysis of a large flow of data, taking advantage of its availability as open sources such as the one handled in social networks, this in addition to looking for a way to structure data such as posts in that networks.

Taking into consideration the influence that has the social networks in the people and the use of this media to promote or organize cyberterrorism, the use of NLP as support on cybersecurity prevention. It has been considered as a way to automate the analysis of a large flow of data, taking advantage of its availability as open sources such as the one handled in social networks, in addition to looking for a way to structure data such as posts in that networks.

In this regard, an architecture that takes advantage of the scope of NLP tools, that extracts information from open sources, mainly thought about social networks and that integrates three NLP models was proposed in this work. This proposal was tested with a prototype and using a dataset extracted from Twitter of a real and recent scenario, this demonstrated the feasibility of the proposal implementation and the scope of the proposed models. Finally, the results prove the value of the analysis of this data for the prevention of cyberterrorism.

As future work, generate experiments using another source of data, like forums or other social networks can be explored to compare results and generate improvement proposals, also an experiment that considers a long period can be set to know the performance or the model if the studied data does not come from a scenario where the generation of terror is more likely.

Another work that contributes to the construction of a cyberterrorism prevention tool could include the integration to the proposal of a model that can predict the intention of the analyzed text. These kinds of models are commonly used to understand the intention of the interlocutor of a conversational model to generate a response, in this case, is possible to study what contribution can be achieved with a model of this type.

REFERENCES

[1] Council of Europe. *Explanatory Report to the Convention on Cybercrime.* `https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016800cce5b`. 2001.

[2] Akhilesh Chandra and Melissa J. Snowe. "A taxonomy of cybercrime: Theory and design". In: *International Journal of Accounting Information Systems* 38 (2020). 2019 UW CISA Symposium, p. 100467. ISSN: 1467-0895. DOI: `10.1016/j.accinf.2020.100467`. URL: `https://www.sciencedirect.com/science/article/pii/S1467089520300348`.

[3] João Rafael Gonçalves Evangelista et al. "Systematic literature review to investigate the application of open source intelligence (osint) with artificial intelligence". In: *Journal of Applied Security Research* (2020), pp. 1–25.

[4] Heather J Williams and Ilana Blum. *Defining second generation open source intelligence (OSINT) for the defense enterprise.* Tech. rep. RAND Corporation Santa Monica United States, 2018.

[5] Javier Pastor-Galindo et al. "The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends". In: *IEEE Access* 8 (2020), pp. 10282–10304. DOI: `10.1109/ACCESS.2020.2965257`.

[6] A. Thomas. *Natural Language Processing with Spark NLP: Learning to Understand Text at Scale.* O'Reilly Media, 2020. ISBN: 9781492047766. URL: `https://books.google.com.co/books?id=sJw6zQEACAAJ`.

[7] Leigh Clark et al. "What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents". In: New York, NY, USA: Association for Computing Machinery, 2019, 1–12. ISBN: 9781450359702. URL: `10.1145.3290605.3300705`.

[8] Swati Kumari, Zia Saquib, and Sanjay Pawar. *Machine Learning Approach for Text Classification in Cybercrime.* 2018. DOI: `10.1109/ICCUBEA.2018.8697442`.

[9] C. Sánchez-Rebollo et al. "Detection of Jihadism in Social Networks Using Big Data Techniques Supported by Graphs and Fuzzy Clustering". In: *Hindawi* 2019.1238780 (2019), p. 13. DOI: `10.1155.2019.1238780`.

[10] Ibrahim Aljarah et al. "Intelligent detection of hate speech in Arabic social network: A machine learning approach". In: *Journal of Information Science* 47.4 (2021), pp. 483–501. DOI: `10.1177/0165551520917651`. eprint: `https://doi.org/10.1177/0165551520917651`. URL: `https://doi.org/10.1177/0165551520917651`.

[11] Iván Castillo-Zúñiga et al. "Internet Data Analysis Methodology for Cyberterrorism Vocabulary Detection, Combining Techniques of Big Data Analytics, NLP and Semantic Web". In: *International Journal on Semantic Web and Information Systems* 16 (Jan. 2020), pp. 69–86. DOI: `10.4018/IJSWIS.2020010104`.

[12] C Oleji et al. "Big data Analitic of Boko Haram insurgency attacks menace in nigeria using DynamicK-reference clustering algorithm". In: 7 (Apr. 2020), pp. 1099–1107.

[13] V. N. Uzel, E. Saraç Eşsiz, and S. Ayşe Özel. "Using Fuzzy Sets for Detecting Cyber Terrorism and Extremism in the Text". In: *2018 Innovations in Intelligent Systems and Applications Conference (ASYU).* 2018, pp. 1–4. DOI: `10.1109/ASYU.2018.8554017`.

[14] Sanjeev J Wagh, Manisha S Bhende, and Anuradha D Thakare. *Fundamentals of Data Science.* Chapman and Hall/CRC, 2021, p. 14.

[15] Sepideh Bazzaz Abkenar et al. "Big data analytics meets social media: A systematic review of techniques, open issues, and future directions". In: *Telematics and Informatics* 57 (2021), p. 101517. ISSN: 0736-5853. DOI: 10.1016/j.tele.2020.101517. URL: https://www.sciencedirect.com/science/article/pii/S0736585320301763.

[16] Amir Karami et al. "Twitter and Research: A Systematic Literature Review Through Text Mining". In: *IEEE Access* 8 (2020), pp. 67698–67717. DOI: 10.1109/ACCESS.2020.2983656.

[17] Yu Ouyang and Richard W Waterman. *Trump, Twitter, and the American democracy: Political communication in the digital age.* Springer Nature, 2020.

[18] Lennart Ante. "How Elon Musk's Twitter Activity Moves Cryptocurrency Markets". In: *Available at SSRN 3778844* (2021).

[19] Ankit and Nabizath Saleena. "An Ensemble Classification System for Twitter Sentiment Analysis". In: *Procedia Computer Science* 132 (2018). International Conference on Computational Intelligence and Data Science, pp. 937–946. ISSN: 1877-0509. DOI: 10.1016/j.procs.2018.05.109. URL: https://www.sciencedirect.com/science/article/pii/S187705091830841X.

[20] A. Poornima and K. Sathiya Priya. "A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques". In: *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. 2020, pp. 493–496. DOI: 10.1109/ICACCS48705.2020.9074312.

[21] Sheresh Zahoor and Rajesh Rohilla. "Twitter Sentiment Analysis Using Lexical or Rule Based Approach: A Case Study". In: *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. 2020, pp. 537–542. DOI: 10.1109/ICRITO48877.2020.9197910.

[22] Beakcheol Jang et al. "Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism". In: *Applied Sciences* 10.17 (2020). ISSN: 2076-3417. DOI: 10.3390/app10175841. URL: https://www.mdpi.com/2076-3417/10/17/5841.

[23] Ashish Vaswani et al. "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[24] Jaekeol Choi and Sang-Woong Lee. "Improving FastText with inverse document frequency of subwords". In: *Pattern Recognition Letters* 133 (2020), pp. 165–172. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2020.03.003. URL: https://www.sciencedirect.com/science/article/pii/S0167865520300817.

[25] Joel Nothman et al. *Learning multilingual named entity recognition from Wikipedia.* 2017. URL: https://figshare.com/articles/dataset/Learning_multilingual_named_entity_recognition_from_Wikipedia/5462500.