UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Dept. of Computer Systems and Computation

Exploring Instagram Messages on the Right to Abortion: User Profiling and Natural Language Processing Tasks

Master's Thesis

Master's Degree in Artificial Intelligence, Pattern Recognition and Digital Imaging

AUTHOR: López Cuerva, Luis

Tutor: Castro Bleda, María José

Cotutor: Hurtado Oliver, Lluis Felip

External cotutor: IRANZO CABRERA, MARIA

ACADEMIC YEAR: 2022/2023

Resum

Les xarxes socials són un dels canals principals de difusió per als moviments socials, les protestes i l'opinió pública. Entre les xarxes socials més populars està Instagram, una xarxa social enfocada en les imatges, vídeos i publicacions temporals. Aquesta plataforma s'ha convertit en un canal comunicatiu més on la gent expressa la seua opinió sobre una gran varietat de temes. Entre els debats presents en aquesta xarxa social destaca el debat sobre el dret a l'avortament, un debat candent en els últims anys. En aquest treball es proponen dos marcs de treball que combinen tasques de processament del llenguatge natural (NLP) i user profiling a nivell de post. Aquests marcs de treball tenen com a objectiu definir un conjunt de passos per a analitzar debats en Instagram de manera ràpida i efectiva utilitzant diferents tipus de embeddings per a representar les dades i múltiples tècniques de clustering per a categoritzar els posts.

Paraules clau: Processament del Llenguatge Natural, Intel·ligència Artificial, Aprenentatge Automàtic, Aprenentatge profund, xarxes socials, Instagram, entitat de nom, perfil d'usuari

Resumen

Las redes sociales son uno de los canales principales de difusión para los movimientos sociales, las protestas y la opinión pública. Entre las redes sociales más populares está Instagram, una red social enfocada en las imágenes, videos y publicaciones temporales. Esta plataforma se ha convertido en un canal comunicativo más donde la gente expresa su opinión sobre un gran variedad de temas. Entre los debates presentes en esta red social destaca el debate sobre el derecho al aborto, un debate candente en los últimos años. En este trabajo se proponen dos marcos de trabajo que combinan tareas de procesamiento del lenguaje natural (NLP) y user profiling a nivel de post. Estos marcos de trabajo tienen como objetivo definir un conjunto de pasos para analizar debates en Instagram de forma rápida y efectiva utilizando diferentes tipos de embeddings para representar los datos y múltiples técnicas de clustering para categorizar los posts.

Palabras clave: Procesamiento del Lenguaje Natural, Inteligencia Artificial, Aprendizaje Automático, Aprendizaje profundo, redes sociales, Instagram, entidad de nombre, perfil de usuario

Abstract

Social networks are one of the main dissemination channels for social movements, protests, and public opinion. Among the most popular social networks is Instagram, which focuses on images, videos, and temporary posts. This platform has become another communication channel where people express their opinions on a wide variety of topics. Among the debates present on this social network, the debate on the right to abortion stands out, a hot debate in recent years. In this work, two frameworks are proposed that combine natural language processing (NLP) tasks and user profiling at the post level. These frameworks aim to define a set of steps to analyze debates on Instagram quickly and effectively using different types of embeddings to represent the data and multiple clustering techniques to categorize posts.

Keywords: Natural Language Processing, Artificial Intelligence, Machine Learning, Deep learning, social media, Instagram, name entity, user profiling

Contents

Appendix

List of Figures

List of Tables

Acknowledgments

Thanks are due to ValgrAI - Valencian Graduate School and Research Network for Artificial Intelligence and to the Generalitat Valenciana for the financial support for the realization of the master's program.

Preface

This final master's thesis arose from a call from Maria José in which she told me about a possible project in collaboration with the Universitat de València. I said yes without being clear about the scope of that project or what my tasks would be. After this first contact, there was a meeting with my two tutors in which I understood both my role and the scope of what had been proposed to me. At that moment, I started a collaboration in which I faced tasks that I had never done before, and I took up again a field of study that I had discovered that I loved during my final degree project: Natural Language Processing. This step was followed by a meeting with several journalists, including María Iranzo Cabrera, Coordinator of the Degree in Journalism at the Universitat de València, and we started working on the project of which my master's final paper would be a part: to analyze the right to abortion on Instagram, to analyze how disinformation spreads in such a current debate and to generate a set of expert tools that allow quick analysis in public, international and full of socio-political edges debate. However, if I have learned anything during my time at the university, it is that coordination between multidisciplinary teams is not an easy task, especially when there are members from different universities. This has led to having to extend the dates initially proposed, although the university academic calendar does not move in time. For this reason, I consider this final master's thesis as the first step of an extremely interesting collaboration in which I can work again with a highly topical subject and a set of cutting-edge artificial intelligence tools whose next step will be the publication of a scientific journal of the results obtained in the work that begins here.

CHAPTER 1
Introduction

1.1 Motivation

Social networking services (SNSs) have been acknowledged as the key channel for protests and social movements [1]. Among the most popular social networks there is Instagram. Instagram is a social network focused on images, comments, and temporal publications. These characteristics make it possible for Instagram users to communicate with others in an easy, fast, and diverse way. These characteristics facilitate the uprising of hashtag activism. Hashtag movement, or hashtag activism, refers to actively utilizing the hashtag function of SNSs for social change [2]. Since it can facilitate spreading awareness and information on a social issue, hashtags can be a helpful tool for activists who struggle for social changes [3]. It is also beneficial for ordinary people as they can easily share their stories, which can be both personal and political, with other people who have similar viewpoints and express support for social movements. Combining the hashtag movement and the personal perspective allows Instagram users to create common narratives between publishers and followers to impact public opinion about a controversial topic.

As with other controversial social issues, abortion has been a significant topic discussed on SNSs, particularly in the context of Instagram. The demand for abortion rights has been a historical struggle for the feminist movement. It has become a booming social movement in recent times, especially in countries such as Mexico, Argentina, Malta, and the United States [4, 5, 6]. However, trends in the abortion rights debate depend on the country where the debate is situated. While in the United States, there is a conservative tendency to make access to abortion more complicated, in Argentina and Mexico, the feminist movement has succeeded in achieving the right to abortion. In 2020, Argentina became the third country in Latin America to provide abortion rights, after Cuba and Uruguay (Mexico had guaranteed this right in 2007, but only in Mexico City and in Oaxaca in 2019, then six other states in 2021-2022) [7]. Meanwhile, in The United States in 2022, the Supreme Court repealed the constitutional right to abortion, reversing a 50-year-old precedent. It was the 1973 “Roe vs. Wade” judgment that expanded women’s access to abortion, which contributed to developing criteria for “legal abortion” in the United States [8].

1.2 Objectives

The work reported in this project consists of four complementary objectives. The first is the creation and analysis of a multimodal dataset on the abortion rights debate; the second is the research, analysis, and use of various natural language processing tools to analyze comments on posts; the third is the research, analysis, and use of various multimodal tools to analyze comments on posts; and the fourth is the definition of a framework for the creation of user profiles in the context of the abortion rights debate on Instagram.

To evaluate the degree of fulfillment of these objectives, the rubric shown in Table 1.1 is proposed, which will be used at the end of the work to see to what extent the objectives have been fulfilled.

Objectives	**Objective not accomplished**	**Objective insufficiently accomplished**	**Objective sufficiently accomplished**	**Objective fully accomplished**
Creation of a dataset	The dataset has not been created.	The dataset only contains publications captions.	The dataset contains publications and comments information.	The dataset contains information about the publications, comments, and metadata.
Definition and use of a user profiling framework at post level	No user profiling techniques have been used.	An external user profiling framework has been used.	It has been proposed a user profiling framework.	Various user profiling frameworks have been defined.
A study of embedding tools to represent the information	Embedding tools have not been used.	Text embedding tools have been used.	Image embedding tools have been used.	Multimodal embedding tools have been used.
Study of NLP tasks over the dataset	No NLP tasks have been addressed	The languages present in the dataset have been studied.	The polarity present in the dataset has been evaluated.	The named entities present in the dataset have been analyzed.
Paper publication	The paper has not been addressed	Pre-publication steps have been carried out	A first draft has been created.	The paper has been published

Table 1.1: Objective evaluation rubric.

1.3 Memory structure

This paper is divided into nine chapters and the bibliography. The specific chapters are:

1. Introduction. It presents this work and its objectives.
2. State of the art. It exposes the current status of the various studies related to this work.
3. Dataset. It describes how the dataset was created and a statistical analysis of the dataset.
4. User profiling. It defines two frameworks for user profiling at the post level.
5. NLP Analysis. It presents dataset analysis in the sentiment analysis tasks and named entities recognition.
6. Embeddings. It presents the techniques used to create embeddings of the data.
7. Clustering. It presents the clustering techniques used during experimentation and the metrics used to evaluate the quality of the created clusters.
8. Experimentation. Details the set of experiments carried out.
9. Conclusions and ongoing work. It describes in detail the conclusions obtained and ongoing work that would finish on the research carried out.

CHAPTER 2

State of the art

This chapter presents the current state of the art of the multiple tasks addressed throughout the thesis.

2.1 Natural Language Processing

Since 2020, Natural Language Processing tasks have become increasingly important since there is an increasing amount of unstructured text that needs to be analyzed in order to be able to perform tasks such as filtering spam messages, analyzing the sentiment of a text, topic detection, named entity recognition (NER), and detecting impersonations. The sentiment analysis, also called opinion mining, consists of determining the expressed polarity of a text. It is useful to extract the views of the writer and their moods.

Following [9], an opinion can be defined in terms of a quintuple $(e_i, a_{ij}, o_{ijkl}, h_k, t_l)$ where e_i is the entity, a_{ij} is an aspect related to the entity e_i , h_k is the opinion holder, t_l is the timestamp when the opinion was emitted and $s_i\ j_{kl}$ is the sentiment expressed by the author h_k about the aspect a_{ij} of the entity e_i with timestamp h_k [10]. The sentiment s_{ijkl} can be modeled in different ways. The most common approach consists of using a discrete taxonomy of sentiments: negative, positive, and neutral (that typically means no sentiment expressed). Also, in some works, the neutral class is considered with different meanings, and it is split into two different classes [11]: neutral and none. In these cases, the term neutral refers to the neutralization of positive and negative sentiments (both expressed with the same intensity), while the term none means no sentiment expressed. These discrete classes can be extended to consider different intensities of the sentiment, e.g., strong negative, negative, neutral, positive, and strong positive. Outside of the discrete taxonomy, the sentiment intensities can also be studied in a more fine-grained way than the previous approach by constraining them to some continuous interval. It is convenient to highlight that these taxonomies are oversimplifications of the sentiment analysis task, and most of the works on sentiment analysis work under them [10].

The sentimental analysis task can be carried out at four different levels [12]:

- Document level: The analysis is performed on a whole document, and a single polarity is given to the whole document.

- Sentence level: Each sentence is analyzed and found with a corresponding polarity. This approach is advantageous when a document has a wide range and mixed sentiments associated with it [13].

- Phrase level: The analysis is performed individually in each phrase, which is useful when each phrase contains only one aspect.
- Aspect level: The analysis is performed at the aspect level. Each sentence may contain multiple aspects, each one with a different polarity. This approach allows to carry out a fine-grained analysis of the document.

Much work has been carried out in different areas and at different analysis levels regarding the sentiment analysis task. In education [14, 15] use an aspect level analysis to extract the different aspects that the students used in an open field text questionnaire to extract the most relevant aspects of their needs and feelings about them. In healthcare and social networks, [16] use analysis at a document level to perform sentiment analysis over tweets, generating a single polarity for tweets about Covid 19 vaccine. [17] aggregates the latest research on sentiment analysis applied to public services, and [18] performs opinion mining and sentiment classification based on user behavior at the document level over reviews of online dating services.

Regarding this area, there are mainly three approximations to the task of sentiment analysis:

- LSTM, RNN, and variations.
- Transformer-based approaches.
- Classica machine learning and others.

The first approach agglutinates all the techniques derived from the Recurrent Neural Networks (RNN) and Long-Short Term Memory (LSTM) neural networks. This kind of neural network has been broadly used for NLP tasks, especially for sentiment analysis, as it is capable of capturing long-range dependencies and handling sequential data. These are the reasons why this kind of neural networks were applied successfully in [19], [20], or [21]. In [19], multiple deep learning, machine learning, and classical algorithms are compared in the task to classify the polarity present in movie reviews. The technique that acquired the best results is a stacked-BLSTM neural network. This model is composed of two long-short term memory layers that are stacked; stacking them allows them to accomplish right-to-left and left-to-right dependencies. In [20], a model called Co-LSTM is proposed with an embedding word model trained through backpropagation and the Word2vec algorithm. The Co-LSTM model is formed by a convolutional layer that pools the most important features in the embeddings, an LSTM layer that sequentially analyses the generated vectors from left to right, and another convolutional layer that predicts the actual sentiment. In [21], a Bi-LSTM Self attention-based CNN (BAC) model is used together with a word vector matrix as a representation of the text. The self-attention mechanism creates a context vector for each word, which reflects the internal spatial relation between each word and the remaining other words, allowing Bi-LSTM Self attention-based CNN to achieve the best results.

The second approach groups the techniques that employ the transformer architecture for the sentiment analysis task. Nowadays, transformers are the most used technique as they stand out in their capability to handle long-range dependencies across the text and weigh their importance due to their self-attention mechanism. Also, currently, plenty of pre-trained transformer models can be fine-tuned for a vast number of different tasks. In [22], BERT architecture is presented. BERT stands for Bidirectional Encoder Representations from Transformers. It is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both the left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering, sentiment analysis, and language inference, without substantial task-specific architecture modifications. One example of this process of fine-tuning can be found in [23], where TWilBert is presented. TWilBert is a pre-trained deep bidirectional transformer for Spanish Twitter. This specialization consists of training a BERT model from scratch to obtain coherent contextualized embeddings of Spanish tweets. In order to learn inter-sentence coherence, they propose Reply Order Prediction (ROP), an adaptation of the NSP signal to Twitter conversations. [24] presents a configuration of BERT called Roberta for a Robustly optimized BERT approach. This configuration improves the model's performance by training the model longer, with bigger batches over more data, removing the next sentence prediction objective, training on longer sequences, and dynamically changing the masking pattern applied to the training data. [25] presents XLM-R, a Transformer based masked language model on one hundred languages, using more than two terabytes of filtered CommonCrawl data, and it provides substantial gains over previous multilingual models like mBERT [26] and XLM [27] on classification, sequence labeling and question answering. In [28], an XLM-T model fine-tuned on a set of unified sentiment analysis Twitter datasets in eight different languages is presented. The authors also compared the Twitter-based multilingual language model with a standard multilingual language model trained on general-domain corpora. Finally, they released the multilingual language model along with starting and evaluation code to facilitate research in Twitter at a multilingual scale (over thirty languages used for training data).

The third approach contains all the classical machine learning and other techniques that do not fit into the previous category. Examples of this group are the techniques presented in [29, 30, 31]. In [29], a k-nearest neighbors (KNN) classifier, in conjunction with a lexicon, is employed to evaluate Iraqi tourism firms based on extracting sentiments from Iraqi dialect reviews. In [30], a workflow for aspect extraction based on rules is proposed. The general process consists of four different phases. In the first one, the data is preprocessed, splitting the texts into sentences. The second phase involves extracting the dependency relations and tagging the different POS in the sentences. In the third phase, the IOWA (Improved Whale Optimization Algorithm) algorithm selects the optimal combination subset of rules from the complete set. Finally, a correction phase is applied to save the approved aspects as final while discarding the incorrect ones. In [31], LSTM (long shot term memory) layers together with a GCN (Graph convolutional network) are proposed to capture the potential sentiment dependencies of the contextual words. The capture is done in two phases; in the first one, LSTM (long shot term memory) layers are used to learn contextual representations of the text. In the second phase, GCN layers, which take the hidden contextual representations of the sentence and corresponding affective enhanced graph as input, are used to capture the potential sentiment dependencies of the contextual words. Afterward, the representations derived from these two components are combined to extract the significant sentiment dependencies with respect to the specific aspect. Here, it is different from most previous graph-based models, which only focus on the syntactical information of the sentence.

Named Entity Recognition (NER) is a natural language processing (NLP) technique that involves identifying and classifying named entities in text into predefined categories. Named entities are words or phrases that refer to specific types of entities, such as names of persons, organizations, locations, dates, percentages, monetary values, and more. They usually carry key information in a sentence, which serve as important targets for most language processing systems. Accurate named entity recognition can be used as a useful source of information for different NLP applications [32]. Mmst NER systems use three intuitive classes of person (PER), location (LOC), organization (ORG) along with the loosely defined miscellaneous(MIS) class[32].

According to [33], there are mainly two approaches for the task named entity recognition:

- Flat named entity recognition: aims to identify named entities where a named entity consists of contiguous tokens and the named entities do not overlap.
- Nested named entity recognition: aims to identify named entities when an entity could contain other entities or be a part of other entities. For example, the entity "the Valencia Zoo" contains an inner entity, i.e., "Valencia".

For the first approach, the most common method is to use sequence tagging techniques with a sequence tag scheme, which allows the model to classify individual tokens (i.e., words) and some consecutive tokens with the same label are combined to identify named entities [33].

For the second approach, multiple neural models have been proposed. For example, in [34], the Layered-BiLSTM-CRF model is presented. This model is composed of multiple bidirectional layer stacks and a CRF (Conditional Random Field) layer, so it is able to capture context representation of input sequences and globally decode predicted labels at a flat NER layer without relying on feature engineering. In [35], the authors use both word embeddings and character embeddings as input, feed the output into a BLSTM, and finally, to an affine classifier to detect the named entities. To encode a word, they used Bert$_{Large}$ and fast-Text embeddings [36], and for BERT, they followed the recipe of [37] to obtain the context-dependent embeddings for a target token with 64 surrounding tokens each side.

2.2 Computer vision models

Transformer-based models have become the state of the art for a wide range of tasks from different research areas. They were initially proposed for NLP tasks where they became state of art for a large number of tasks such as named entity recognition, sentiment analysis, summarization, or text generation. Inspired by the major success of transformer architectures in the field of NLP, researchers have recently applied Transformers to computer vision (CV) tasks and have explored whether similar models can learn valuable representations for images [38].

Among all the models based on transformers for computer vision, the following stand out:

- ViT: Vision Transformers.
- Swin.
- Transformer-Based Set Prediction for Detection.
- VideoMAE.

Vision Transformer (ViT) [39] is a pure transformer directly applied to the sequences of image patches for image classification tasks. It follows the Transformer's original design as much as possible. ViT yields modest results when trained on mid-sized datasets such as ImageNet, achieving accuracies of a few percentage points below ResNets of comparable size. Because transformers lack some inductive biases inherent to CNNs–such as translation equivariance and locality–they do not generalize well when trained on insufficient amounts of data. However, the authors found that training the models on large datasets (14 million to 300 million images) surpassed inductive bias. When pre-trained at a sufficient scale, transformers achieve excellent results on tasks with fewer data points. For example, when pre-trained on the JFT-300M dataset, ViT approached or even exceeded state of the art performance on multiple image recognition benchmarks. Specifically, it reached an accuracy of 88.36% on ImageNet and 77.16% on the VTAB suite of 19 tasks [38].

Shifted windows transformer (Swin) [40] is a hierarchical Transformer whose representation is computed with Shifted windows. The shifted windowing scheme brings greater efficiency by limiting self-attention computation to non-overlapping local windows while allowing cross-window connection. This model has been tested on an extensive, broad range of vision tasks, including image classification (87.3 top-1 accuracy on ImageNet-1K) and dense prediction tasks such as object detection (58.7 box AP and 51.1 masks AP on COCO testdev) and semantic segmentation (53.5 mIoU on ADE20K val).

Transformer-Based Set Prediction for Detection (DETR) [41] is a simple and fully end-to-end object detector that treats the object detection task as an intuitive set prediction problem, eliminating traditional hand-crafted components such as anchor generation and non-maximum suppression (NMS) post-processing. DETR is a new design for the object detection framework based on Transformer and empowers the community to develop fully end-to-end detectors. However, the vanilla DETR poses several challenges, specifically, a more extended training schedule and poor performance for small objects [38].

VideoMAE [42] is a masked autoencoder, a data-efficient learner for self-supervised video pretraining. VideoMAE introduces two critical designs of extremely high masking ratio and tube masking strategy to make the video reconstruction task more challenging. This more complex task would encourage VideoMAE to learn more representative features and relieve the information leakage issue. This design allows VideoMAE with the vanilla ViT backbone to achieve 87.4% on Kinects-400, 75.4% on SomethingSomething V2, 91.3% on UCF101, and 62.6% on HMDB51.

2.3 Multimodal models

Multimodal learning tasks are those that merge data of different typologies. This type of task is booming thanks to advances in natural language processing and computer vision tasks. This type of task presents many different approaches depending on the data to be fused.

In [43], ALIGN is presented. ALIGN is a simple dual-encoder architecture that learns to align visual and language representations of the image and text pairs using a contrastive loss's strong performance when transferred to classification tasks such as ImageNet and VTAB. The authors of [44] introduce an image-conditioned masked language modeling (ICMLM), a proxy task to learn visual representations over image-caption pairs. ICMLM consists of predicting masked words in captions by relying on visual cues. To tackle this task, multiple hybrid models, with dedicated visual and textual encoders, and we show that the visual representations learned as a by-product of solving this task transfer well to a variety of target tasks. In [45], ConVIRT is proposed. ConVIRT is an alternative unsupervised strategy to learn medical visual representations by exploiting naturally occurring paired descriptive text. The proposed method of pretraining medical image encoders with the paired text data via a bidirectional contrastive objective between the two modalities is domain-agnostic and requires no additional expert input.

CHAPTER 3

Dataset

To do the current work, we have created a dataset of images, posts, and comments from Instagram. In the current chapter, the requirements of the dataset, the process of creating it, and an analysis of the dataset will be exposed.

3.1 Requirements

In order to explore the Instagram Messages on the right to abortion, we have scrapped Instagram to retrieve posts, comments, and media to create the dataset. To do this scrapping, we have used Instaloader [46] tool to scrap data from six different hashtags: #noalaborto, #salvemoslasdosvidas, #sialavida, #mareaverde, #quesealey, and #provida. We have chosen these hashtags to scrap to obtain publications from multiple social perspectives as the hashtags #noalaborto, #salvemoslasdosvidas, #sialavida, and #provida are usually hashtags used by anti-abortion movement, and the hashtags #mareaverde and #quesealey are usually used by the pro-abortion movement. It is essential to highlight that obtaining the dataset is scrapping Instagram's explorer. Instagram shows the publications on its explorer, intending to help the user discover new things [47]. To do this, they consider some information about the logged account, the post it will show to the account, and the interaction between the logged account and the new post. About the logged account, Instagram considers what posts have been commented on, liked, and saved. About the new post, it considers how many and how quickly other people are liking, commenting, sharing, and saving a post and information about the person who posted the post, as well as how many times people have interacted with that person in the past few weeks. The relationship between the new post and the account takes into account the history of interaction between the publisher of the new post and the account that will see the new post.

To minimize the effect of Instagram's algorithm, all the scrapping has been carried out from a new account created only for this task, with no prior past experiences in the interaction between the new account and any user and with no likes, saved posts, or comments. However, during this scrap phase, there will be searches of the different hashtags shown before, so Instagram may consider this information while scrapping to show some publications and hide others.

Despite these efforts, the number of posts retrieved by hashtags is very uneven. We are recovering a much smaller number of posts on pro-choice hashtags than on anti-choice hashtags. Table 3.1 shows the number of publications retrieved by hashtags. Multiple hashtags are often used in Instagram posts. This happens in many posts retrieved, so the sum of the number of posts per hashtag does not match the total number of posts retrieved.

Hashtags	Number of posts retrieved
#noalaborto	6852
#salvemoslasdosvidas	959
#sialavida	2020
#mareaverde	27
#quesealey	1
#provida	3344

Table 3.1: Number of posts retrieved by hashtag.

3.2 Retrieved information

In order to obtain the dataset, we have scrapped Instagram's explorer. The scrapped is done from a post perspective, where we scrap posts that contain specific hashtags. From each post that contains one of those hashtags, we retrieve the media of the publication, the comments, and the post information. About each post, we retrieve the following information:

Information retrieved at post level
Id of the post
Date
Title of the post
Profile that published the post
Caption of the post
Tagged users
Accessibility caption
Number of comments
Caption hashtags
Caption mentions
Likes
Url of the publication
Location
Sponsor users
Received comments id
Media of the publication

Table 3.2: List of Post Information

From each comment, we retrieve the following:

Information retrieved at comment level
Id of the comment
Id of the parent post or comment
Text of the comment
Profile that published the comment
Date
Received comments id

Table 3.3: List of Comment Information

From each media, we retrieve:

All of this information is retrieved and then stored in a MongoDB database [48].

Information retrieved at media level
Id of the publication where the media appears
Media

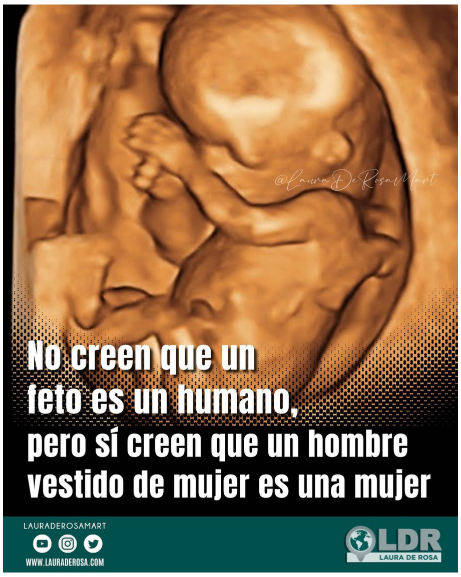
Table 3.4: List of Media Information

3.3 Examples

Posts are the central communicative unit in Instagram, and they are composed of one or more media documents, an image or video (usually an image), and a caption created by the user. In order to illustrate this fact below, we present the five posts with the most interactions, i.e., with the highest combined number of likes and comments. In order to maintain the anonymity of the persons mentioned in the publications, the face of any non-public person appearing in the images is blurred, and mentions to profiles with a @mention are replaced by @mention.

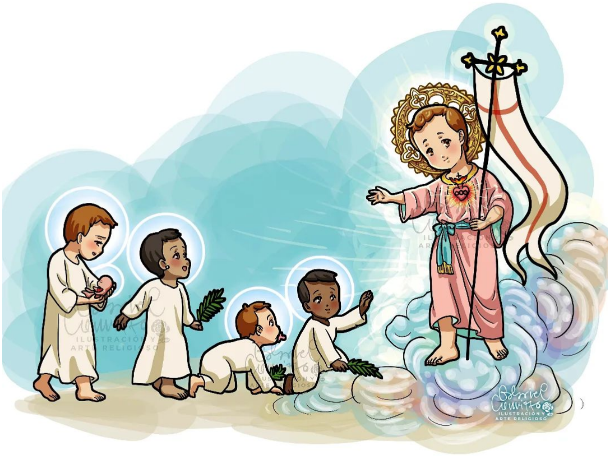
Caption: Me encantó este videito que te comparto ! Ojalá que ayude a tomar conciencia de lo que se está promoviendo desgraciadamente en varios lugares del mundo… Aquí te dejo este testimonio en primera persona. Cuidemos y protejamos la vida de todos ! Comparte este mensaje! Bendiciones! @mention..........#vida #vidahumana #noalaborto #noalabortosialavida #down #downsyndrome #sindromededown

Figure 3.1: Example 1: 8095 interactions. Publicated on 05/02/2023.



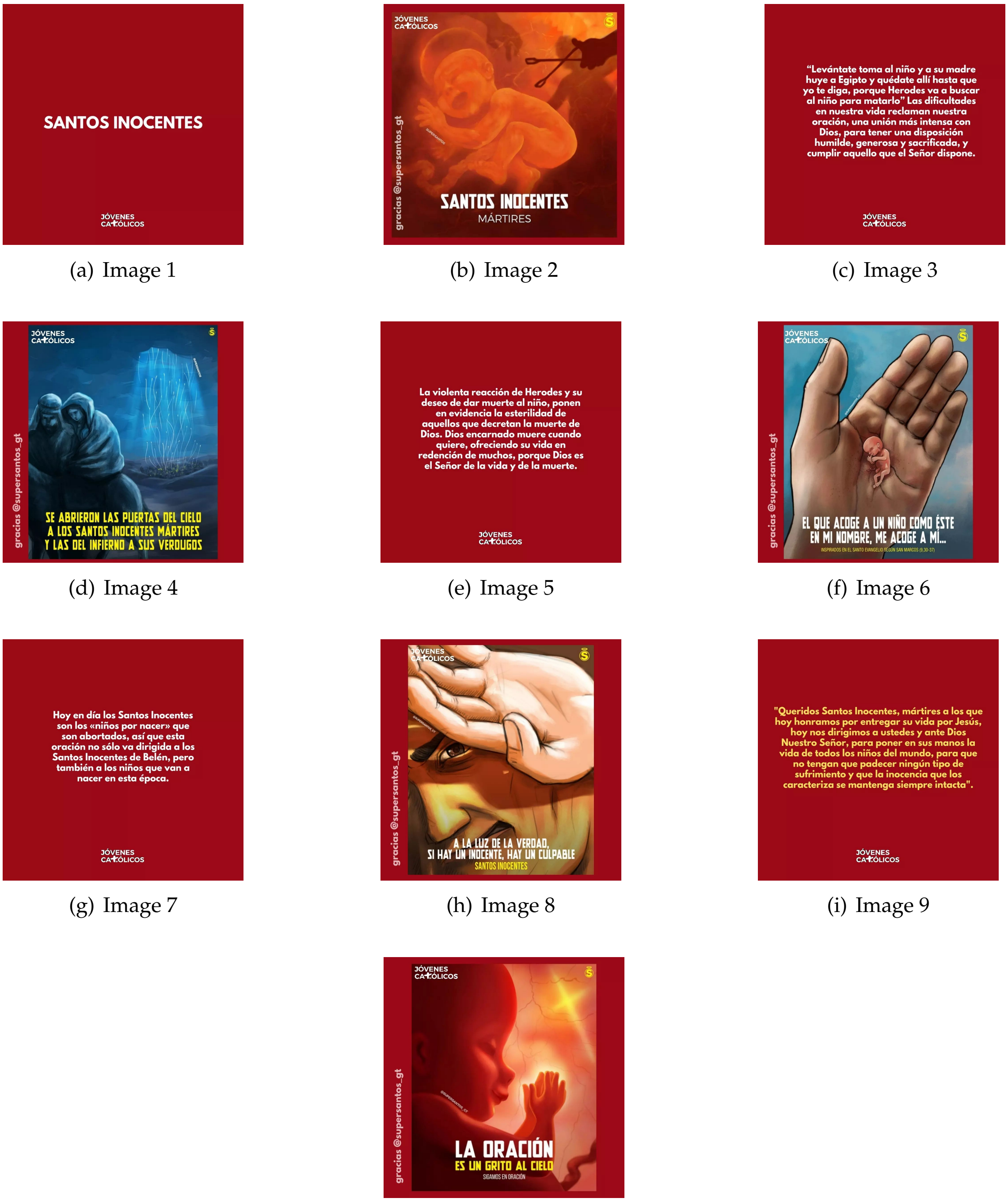
Caption: De coherencia no se van a morir. . . Tampoco de inteligencia.ero en fin, los leoCompártelo (y etiquétame para irte a comentar por tu lado también) ...#LobbyPolitico #PorDetras #Lobby #NoAlAborto#ProVida#Trans ConLosNiñosNo Balenciaga #Trans#Transgenero#Biologia #LeyTrans #IreneMontero#Igualdad#EnLasCompetenciasFemeninasNo #LGTBQRSTUWXYZ #LGTBQ #Mujeres #Mujer#Amigues #Feministas #Patriarcado y la cosa ?#agenda2030 #NiConRojosNiConAzules #FueraElSocialismoDeVenezuela#Venezuela @LauraDeRosaMart #LauraDeRosaMart #LauraDeRosa #LDR

Figure 3.2: Example 2: 5576 interactions. Publicated on 26/05/2023.



Caption: 28 de diciembre: LOS SANTOS INOCENTES (mártires)Los niños Inocentes murieron por Cristo, fueron arrancados del pecho de su madre para ser asesinados: ahora siguen al Cordero sin mancha, cantando: «Gloria a ti, Señor.» (Antífona del Cántico Evangélico de Laudes)Ayer Herodes, que arremetió contra los más pequeños por miedo a perder todo su poder y riqueza a manos de un rey que solo vino a reinar en los corazones. Hoy nuestros legisladores y hermanos de este suelo patrio y de todo el mundo que por miedo a perder su estado de bienestar votan leyes que inventan el derecho a terminar con la vida de los más pequeños e indefensos. Ellos ya gozan en la Gloria de Dios, nosotros roguemos que Él tenga misericordia de aquellas almas que perdieron el camino y han abogado a favor de esta causa y contribuyen directamente a estas muertes.#jesus #niñojesus #babyjesus #meninojesus #gesubambino#santosinocentes#holyinnocents #santosmartires#holymartyrs#martires #martyrs #noalaborto #sialavida #salvemoslas2vidas #iglesiacatolica #dibujo #drawing #ilustracion #illustration #arte #art #ilustraciondigital #digitalillustration #artedigital #digitalart #artereligioso #religiousart

Figure 3.3: Example 3: 5536 interactions. Publicated on 28/12/2022.



(a) Image 1

(b) Image 2

"Levántate toma al niño y a su madre huye a Egipto y quédate allí hasta que yo te diga, porque Herodes va a buscar al niño para matarlo" Las dificultades en nuestra vida reclaman nuestra oración, una unión más intensa con Dios, para tener una disposición humilde, generosa y sacrificada, y cumplir aquello que el Señor dispone.

(c) Image 3

JÓVENES CATÓLICOS
gracias @supersantos_gt
SE ABRIERON LAS PUERTAS DEL CIELO
A LOS SANTOS INOCENTES MÁRTIRES
Y LAS DEL INFIERNO A SUS VERDUGOS

(d) Image 4

La violenta reacción de Herodes y su deseo de dar muerte al niño, ponen en evidencia la esterilidad de aquellos que decretan la muerte de Dios. Dios encarnado muere cuando quiere, ofreciendo su vida en redención de muchos, porque Dios es el Señor de la vida y de la muerte.

(e) Image 5

JÓVENES CATÓLICOS
gracias @supersantos_gt
EL QUE ACOGE A UN NIÑO COMO ÉSTE
EN MI NOMBRE, ME ACOGE A MÍ...
INSPIRADOS EN EL SANTO EVANGELIO SEGÚN SAN MARCOS (9,30-37)

(f) Image 6

Hoy en día los Santos Inocentes son los «niños por nacer» que son abortados, así que esta oración no sólo va dirigida a los Santos Inocentes de Belén, pero también a los niños que van a nacer en esta época.

(g) Image 7

(h) Image 8

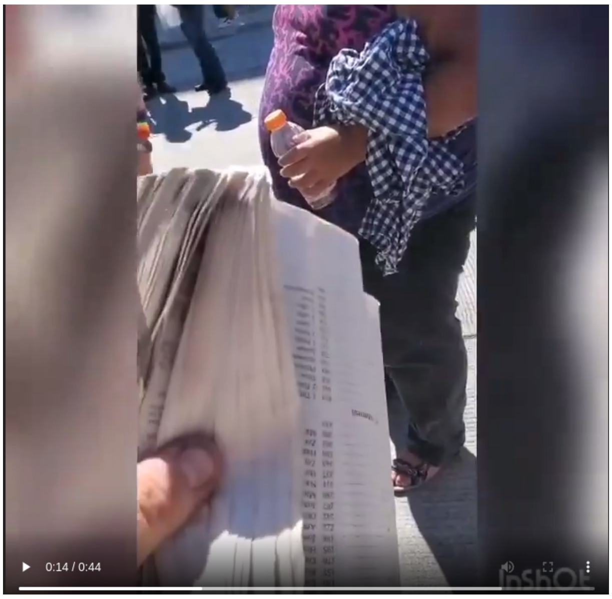
"Queridos Santos Inocentes, mártires a los que hoy honramos por entregar su vida por Jesús, hoy nos dirigimos a ustedes y ante Dios Nuestro Señor, para poner en sus manos la vida de todos los niños del mundo, para que no tengan que padecer ningún tipo de sufrimiento y que la inocencia que los caracteriza se mantenga siempre intacta".

(i) Image 9

(j) Image 10

Caption: "Madre de los niños que no han nacido, ruega por nosotros"..Señor Jesús: por mediación de María, Tu Madre, que te dio a luz con amor, y por intercesión de San José, quien contempló extasiado el Misterio de la Encarnación y se ocupó de Ti tras tu nacimiento, te pido por este pequeño no nacido y que se encuentra en peligro de ser abortado. Te pido que des a los padres de este bebé amor y valor para que le permitan vivir la vida que Tú mismo le has preparado. Amén..Bendito seas, Señor, por este nuevo día. Te alabo por el don de la vida. Al despertar del sueño, te pido especialmente por aquellos que serán trágicamente privados de la vida porque serán abortados. Recíbelos, Señor. Y en tu gran misericordia, guía con tu sabiduría a todas las mujeres embarazas que estén pensando hoy en destruir a los niños que llevan en su seno. Dales la gracia, el valor y la fortaleza para vivir diariamente según tu voluntad. Te lo pido por Cristo, Nuestro Señor, Amén..#rezarhoy #jovenescatolicos #santosinocentes #noalaborto #sialavida #amordeDios #vivirlafe #oracion #testimonios #reflexiones #amistadconjesucristo #fe #alegria #cristianos #givenfaith

Figure 3.4: Example 4: 4270 interactions. Publicated on 28/12/2022.



Caption: Protestantes LGBTQ a favor del aborto le quitan su biblia a cristiano, la pisotean y la echan al inodoro.¡Oh Señor, grande es tu misericordia!#SoyProvida #NoAlAborto

Figure 3.5: Example 5: 3654 interactions. Publicated on 28/06/2022.

3.4 Analysis of posts

We have carried out a previous dataset analysis to know the characteristics of it. In total, we have retrieved 5865 posts, 23208 comments (22775 comments that are direct replies to a post and 433 comments that are replies to another comment), and 10381 media content, which includes photos and videos.

Regarding the posts, we have recovered 5865 posts publicated between 16-06-2020 and 30-05-2023. The posts retrieved received a total of 22775 comments and 397944 likes. 5681 unique hashtags were used a total of 128016 times throughout the publications, and only one publication had the location of the publication indicated.

3.4.1. Statistical analysis

All the details about the post statistics are presented in Table 3.5.

	Total	Maximum	Average	Standard deviation	Median
Likes	397944	7675	67.85	244.80	12
Comments	23208	638	3.88	21.59	0
Interactions	420719	8095	71.73	258.39	13
Hashtags	5681	38	21.82	11.34	29
Mentions	1617	24	0.27	1.34	0

Table 3.5: Posts statistics.

The distributions of all items related to publications show extreme values. Throughout this section, we will study the distributions of each of the elements and analyze possible causes of their characteristics.

First, we will analyze how the number of likes received by the publications is distributed. In Figure 3.6, we have represented on a logarithmic scale this distribution, and the first characteristic fact is that we are in front of a distribution with a very strong mode and an extremely long tail. Most of the publications receive very few likes, and a few publications reach a large number of likes.

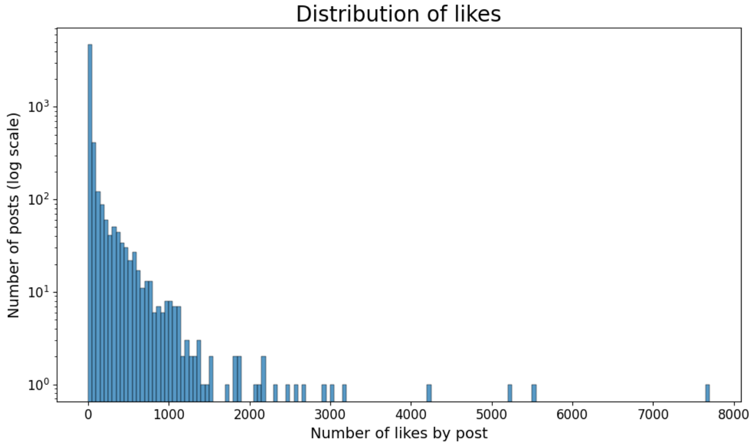


Figure 3.6: Likes distribution.

Secondly, we will proceed to analyze the number of comments received by each publication. This can be seen in Figure 3.7, where we use a logarithmic scale. As with the number of likes per publication, we are faced with a distribution with a very strong mode and a long tail, although in this case, the extreme values are slightly more common.

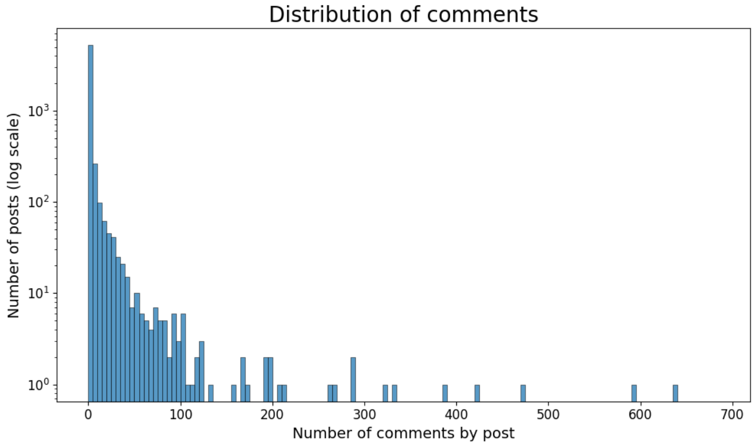


Figure 3.7: Comments distribution.

The third element analyzed is the number of interactions. This has been represented in Figure 3.8 by a logarithmic scale. Once again, we are faced with a distribution of similar characteristics to those observed in Figure 3.6 and Figure 3.7. In this case, it is important to remember that we define the number of interactions of a publication as the sum of the number of likes and the number of comments.

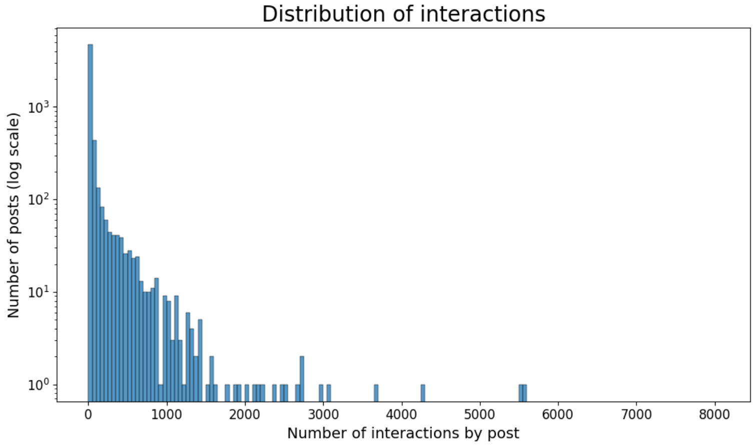


Figure 3.8: Interactions distribution.

Below, we study how the number of hashtags used per publication is distributed. This distribution is represented in the Figure 3.9. Contrary to the previous distributions, here we observe a less sharp distribution where the tail is generated to the left of the mode. Generally, a publication has more than 30 hashtags, but none of the retrieved publications has 40 or more different hashtags.

Finally, we will analyze how the number of mentions used in each publication is distributed. This can be seen in Figure 3.11, which uses a logarithmic scale. In this graph we find again patterns previously seen in Figures 3.6, 3.7, and 3.8, although attenuated.

The first three distributions analyzed are directly related to the reach of the publications, i.e., they are related to how many people see the publication and interact with it. The majority of the publications have a small reach, receiving very few likes and comments, and a minimal part of the publications receive a large number of interactions; the rest of the publications are at an intermediate point in which they reach an average visibility.

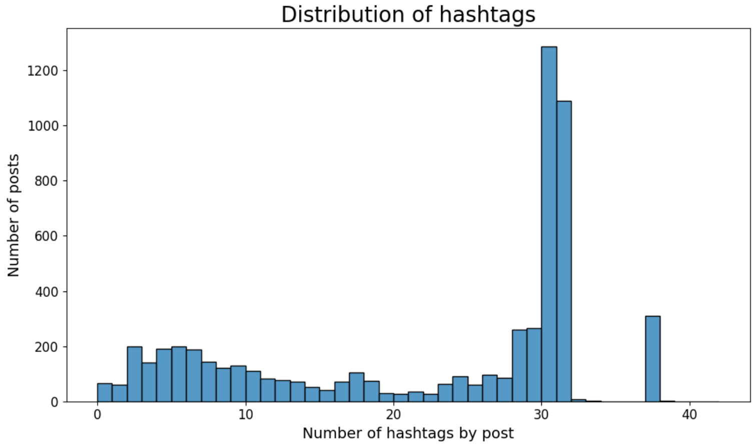


Figure 3.9: Hashtags distribution.

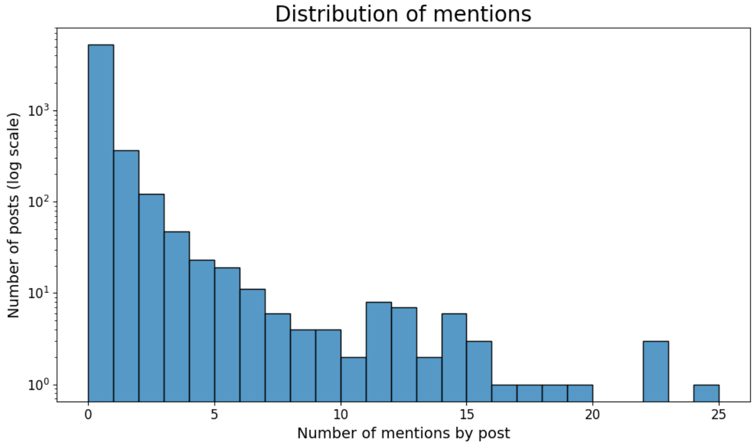


Figure 3.10: Mentions distribution.

Mentions on Instagram are often used in order to increase the visibility of the account so that the posts are seen by a larger number of people. This fact is consistent with the distributions observed previously; most of the publications barely contain mentions of other users, as well as most of the publications have limited visibility and do not receive comments or likes. However, there are a small number of posts that contain a large number of posts, as well as a limited number of posts with a large reach. This effect is due to Instagram's policy when presenting new posts in the feed, where one of the factors to show more users a post is the number of times the account has been interacted with [47].

3.4.2. Dates

In Figure 3.11, we have represented the number of publications recovered depending on the month of publication. Due to the operation of Instagram, mainly recent publications have been recovered, with the largest step in the recovered publications being observed between September 2022 and October 2022. Although we can consider that from October 2022 onwards, we have recovered all types of publications, we must also keep in mind that posts prior to October 2022 are those that Instagram considers of special interest to the account used to do the scrapping (this account was created from scratch for this task), so they probably come from Instagram accounts with a large scope.

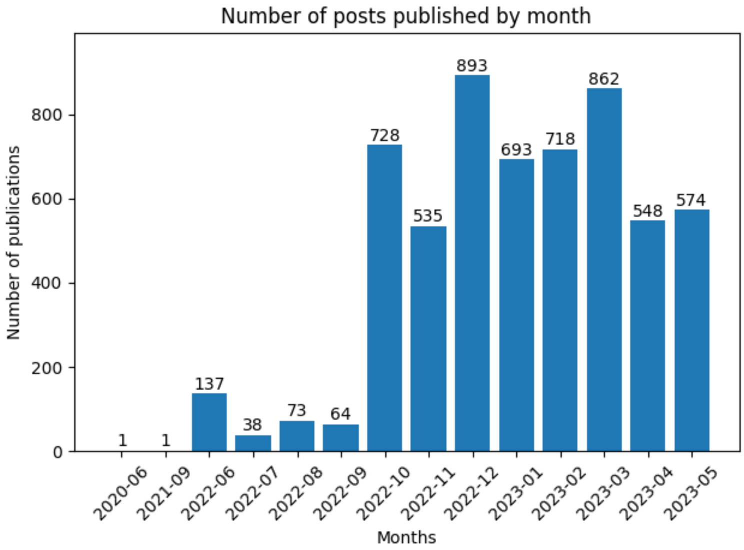


Figure 3.11: Distribution of publication date.

Table 3.6 shows the ten days with the highest number of publications and the events associated with those days. It can be seen that 50% of the dates in the table are directly related to Christianity. The large number of publications on Saint Innocents' Day and the days before and after it stand out within these publications. Concerning the days unrelated to any religion, the presence of March 8, International Women's Day, and the day after International LGBTIQ+ Pride Day can be observed. Finally, the day 09/10/2022 stands out, where there is an increase in the number of advertisements made. This day is the day after massive pro-choice demonstrations in Washington. These protests follow the overturning of the landmark American ruling known as 'Roe v. Wade', in which the U.S. Supreme Court ruled that the U.S. Constitution protects a pregnant woman's freedom to choose to have an abortion without excessive government restrictions.

Date	Num. publications	Associated Event
2022-12-28	234	Holy Innocents Day
2022-12-27	204	Day before Holy Innocents Day
2022-10-09	102	Day after pro-abortion demonstrations in the U.S.[1]
2023-03-25	89	Unborn Child Rights Day
2022-06-28	64	International LGBTIQ+ Pride Day
2023-02-02	62	Candelaria Day[2]
2022-12-29	59	Day after Saints Innocents Day
2022-06-29	52	Day after International LGBTIQ+ Pride Day
2023-03-08	47	International Women's Day
2023-01-26	45	Day of the Missionary Childhood

Table 3.6: Top 10 days with more publications

3.4.3. Most active profiles

In order to look for patterns in the accounts with the highest publication rate, we have represented the 40 accounts with the highest number of publications in Figure 3.12.

Most accounts with the highest number of posts are explicitly created to disseminate information related to abortion rights. More specifically, to disseminate against the right

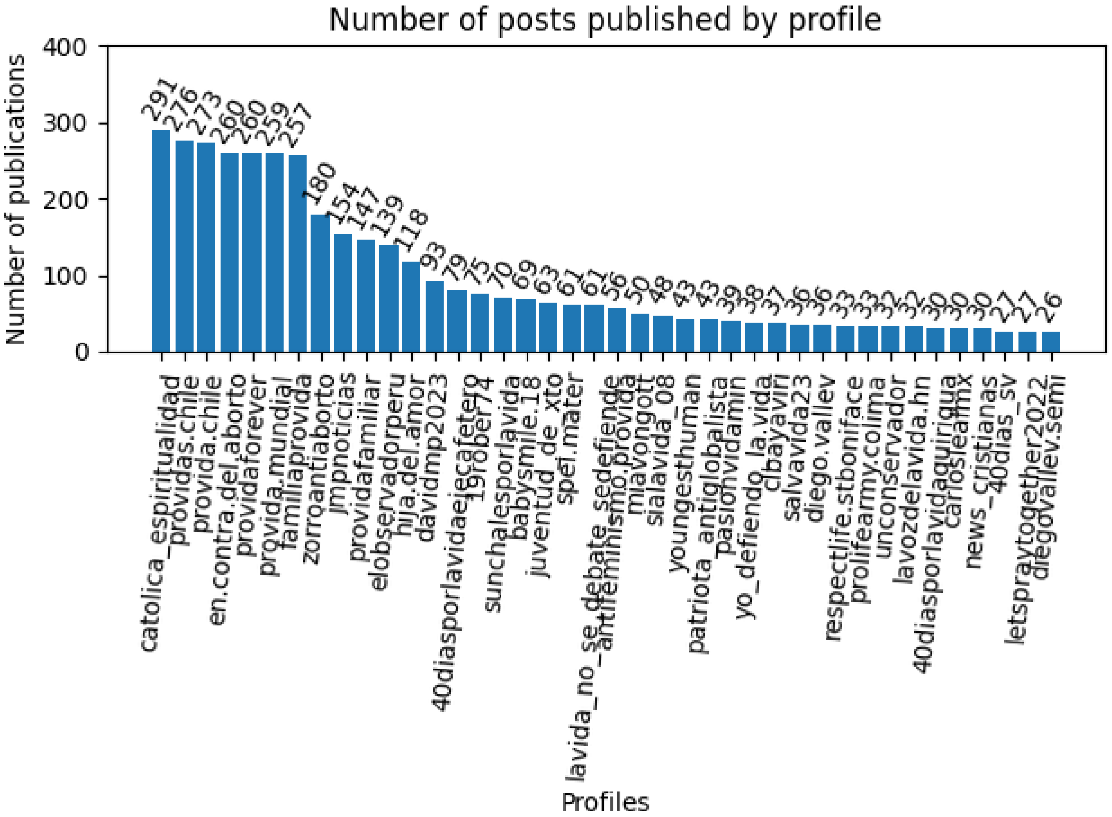


Figure 3.12: Publication distribution.

to decide. Besides this account, the string "provida" is the most commonly used word to identify an account as a disseminator of anti-abortion ideas. Besides this, some of these accounts use the name of a country (Chile, Peru) to identify themselves as members of a particular country and amplify the reach of their message among the inhabitants of that country. It is also noticeable that the account publishing more posts about the right to abortion is called after a religion.

Regarding the distribution of publications by author, the jump in the number of publications between the first seven accounts and the rest, where the largest step in the number of publications takes place, stands out. From this position onwards, the reduction in the number of publications decreases incrementally without abrupt changes.

3.4.4. Languages employed

In the retrieved dataset, we have distinguished the use of 4 languages:

- Spanish: 5427 posts.
- Portuguese: 239 posts.
- Italian: 90 posts.
- English: 81 posts.

In addition to that, there are 28 posts where the caption is empty, so no language is used at all. In Figure 3.13, the % of the use of each language is shown.

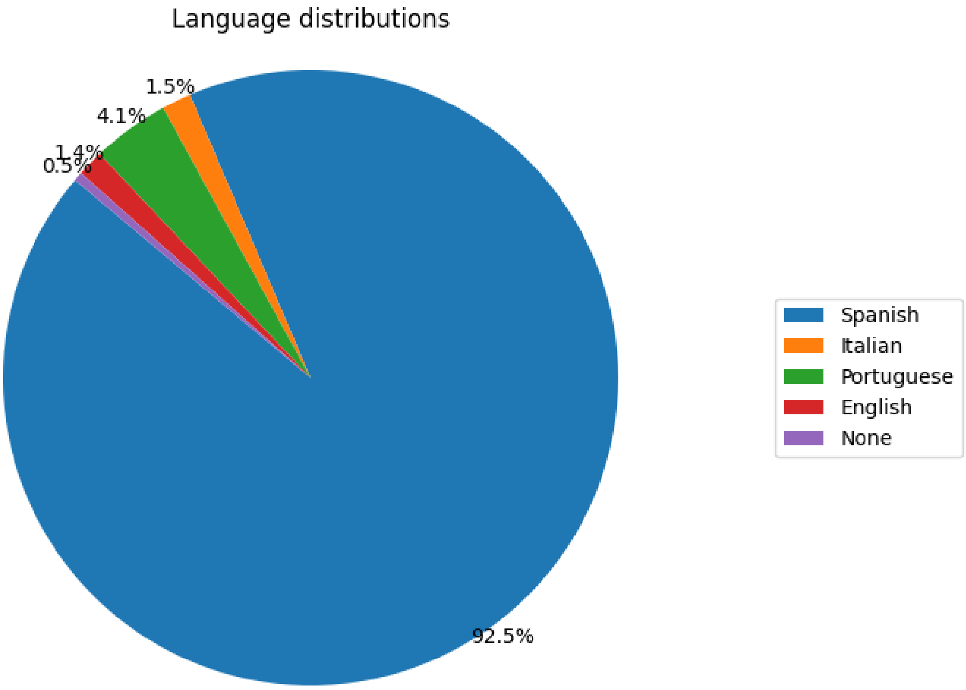


Figure 3.13: Language distribution.

3.5 Analysis of comments

In addition to the publications, the different comments recovered have also been analyzed. Comments are responses to posts or other comments, so they are generally smaller texts than the captions of the posts and do not contain media content.

3.5.1. Dates

First of all we have studied the publication dates of the comments. Figure 3.14 shows how many comments have been posted each month. It is important to note that comments are responses to publications, so they always occur temporarily after publications, and not all publications have the same reach, so an increase in publications does not have to imply a greater number of comments. , since a viral post can have more comments than 100 posts made by profiles with little engagement.

Figure 3.14 shows the temporary irregularity in comments publications. Although it is normal that the most recent publications do not have a large number of comments (since, at the time of making the scrapping, they had not yet received possible comments that they could receive in the days following its publication) this irregularity is also maintained in most previous comments. This fact highlights the existence of viral posts that receive a large number of comments that are combined with posts that hardly receive comments.

3.5.2. Most active profiles

As we have done at the post level, we have studied the profiles with the greatest number of comments. We have represented the forty accounts with the most comments in Figure 3.15.

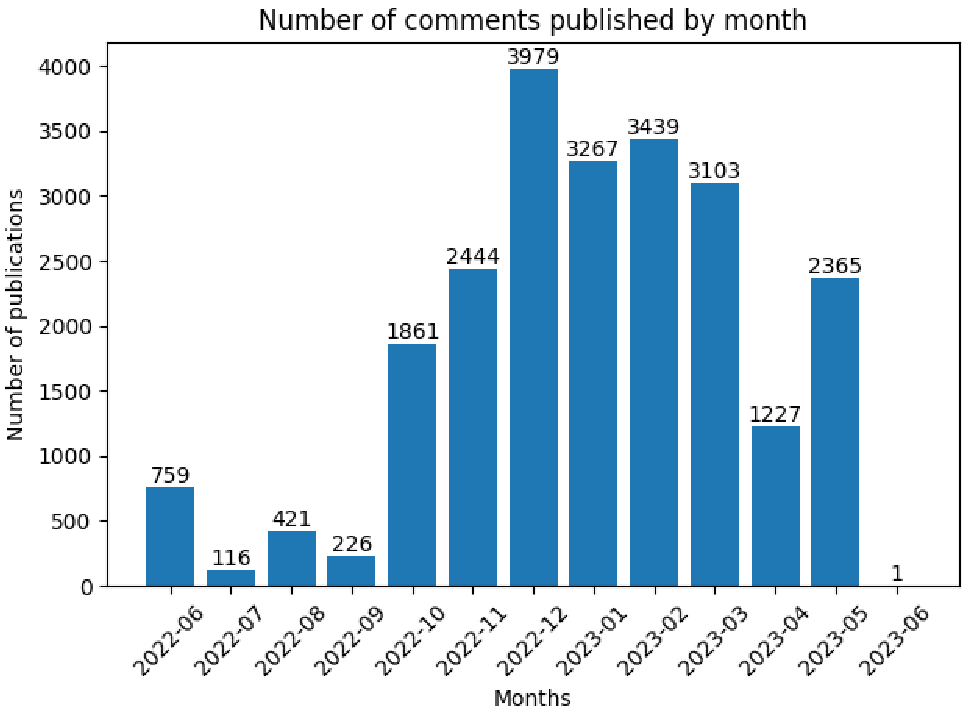


Figure 3.14: Distribution of comments date.

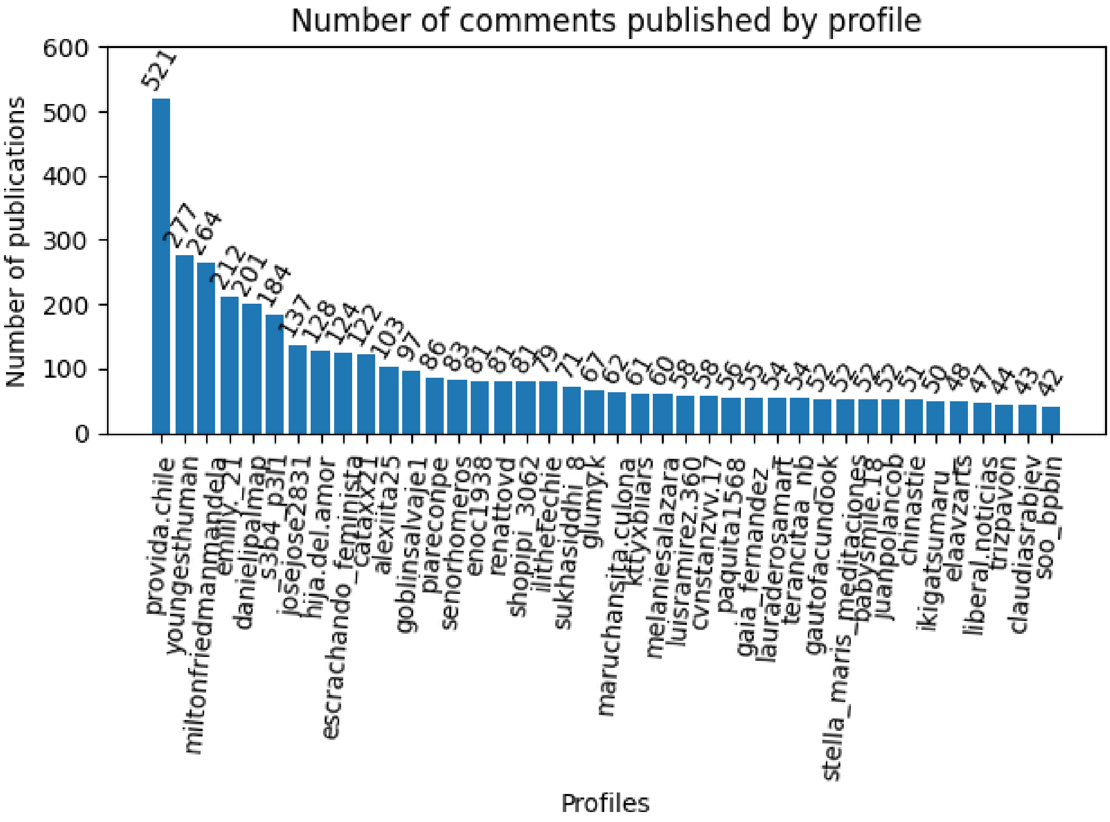


Figure 3.15: Distribution of profiles comments.

Regarding the number of comments per Instagram profile, the "provida.chile" profile stands out for having published a much higher number of comments than the rest of the accounts. Furthermore, this account is the third account with the most post publications, thus being the account that generates the most interactions. This is an account created solely for the purpose of participating in the debate on the right to abortion and spreading ideas against it. From the second onwards there are more comments made, and the decrease in published comments is gradual without any major steps.

CHAPTER 4
User profiling

We will now proceed to define a framework to automatically classify Instagram posts that occur in the context of a debate and the metrics to evaluate the quality of the classification. To this end, we will first proceed to create the framework and then to define the metrics that we will use to measure the quality of the different methods. This thesis proposes two frameworks for the creation of user profiles:

- An extended framework that allows further study of the set of publications to be analyzed generates post categories and classifies posts into those categories.
- A reduced framework that allows the generation of post categories and classification of the posts.

4.1 Extended framework

The extended framework aims to define a set of steps to perform a complete analysis of any discussion on Instagram while minimizing the amount of resources required. Within this framework, two complementary paths are proposed to achieve a high understanding of the discussion, its characteristics, and the type of posts published:

- Analysis of publications: This way of analysis aims to discover what types of publications are made, what are the characteristics of each type of publication and to propose an automatic classification of new publications in the types of publications found.
- Analysis of the content of the publications: In this analysis, the content of the publications is studied, with particular emphasis on the analysis of the polarity present in the publications and the recognition of the different named entities that appear.

Figure 4.1 represents the 6 phases that comprise this framework and the order in which they should be carried out. The phases are as follows:

- Retrieval of posts: Retrieval of a set of posts dealing with the debate to be analyzed. It is recommended to select a set of hashtags dealing with different perspectives of the debate and retrieve posts containing both hashtags. At a minimum, the multimedia content contained in the post and its caption should be retrieved.
- Creation of the embeddings: This is the phase in which the embeddings that represent the information present in the publications are created. These embeddings can

be created in many different ways; however, due to the characteristics of Instagram posts, there are three approaches of special interest: embeddings created from the captions of the posts, embeddings created from the media content of the posts, and embeddings created from text and media. In Section 6, we provide more information about these types of embeddings, and in Section 8 we evaluate them in order to recommend the most useful embeddings for the task addressed.

- Clustering process: Once the different representations of the publications have been obtained, the different types of publications and the characteristics of each type of publication must be found. Since, in this problem, we do not have a predefined set of publication types, clustering techniques must be used to discover the types of publications from the data. In 10

- Detection of the language of the posts: In social networks, languages are usually mixed to the point that a post may contain different languages. This effect also happens in hashtags containing posts from multiple countries and different languages. In order to choose the models that best fit the retrieved posts' characteristics, the first step is to analyze the languages present in the posts. This analysis has been carried out as part of the dataset analysis and can be found in Section3.4.4

- Sentiment analysis: In order to characterize the type of debate that is taking place, it is essential to analyze the polarity of the publications present in the dataset. The analysis can be found in Section 5.

- Named entity recognition: Knowing the different named entities present in the dataset is as essential as knowing the polarity of the different posts. These entities provide much extra information since they allow us to elucidate which geographical areas have the most significant influence on the debate and which organizations and personalities are the most present in the debate. An example of this kind of analysis can be found in Section 5.

One of the main advantages of this framework is the parallelization of the analysis of the content of publications with the analysis of the type of publications. Parallelization speeds up the analysis process, thus reducing the associated costs. At the same time, the framework is agnostic to the artificial intelligence models used. It can be used with any state-of-the-art model and allows both monolingual and multilingual analysis, depending on the models applied.

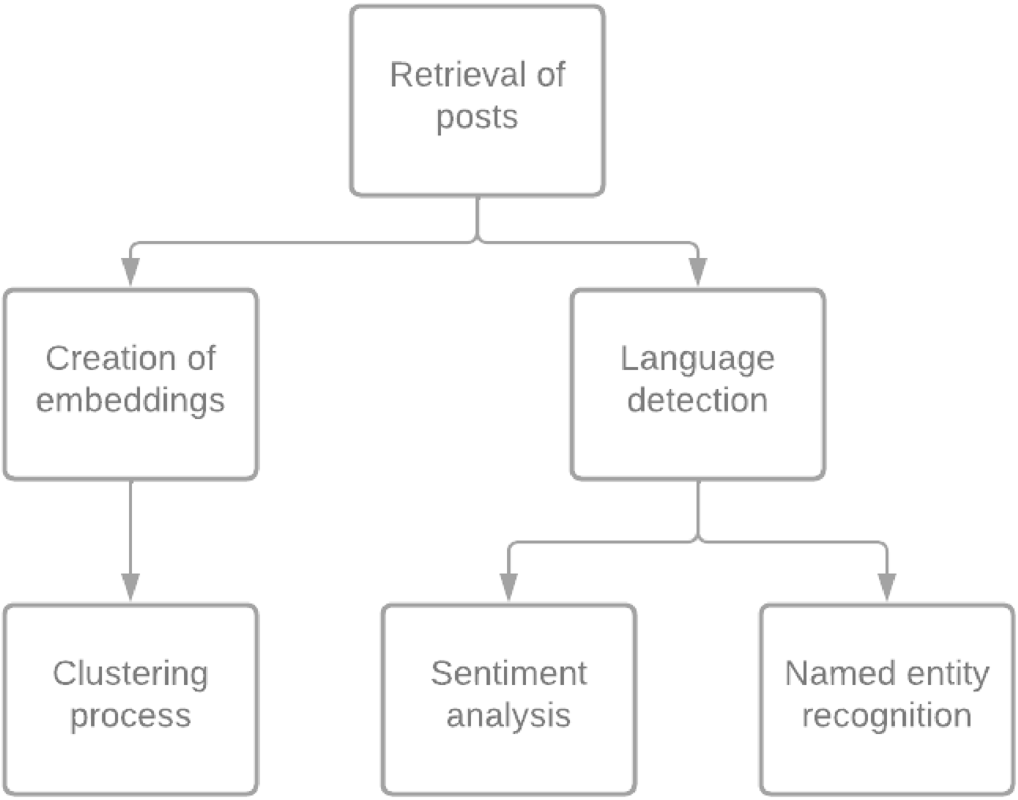


Figure 4.1: Scheme of the extended approach to user profiling.

4.2 Reduced framework

As an alternative to the extended process, we also propose a reduced framework, which allows an extremely fast categorization of the types of publications. However, this reduced framework performs a minor analysis of the dataset. This framework is especially recommended when there is a very short time to analyze a debate on Instagram, either because it is needed to analyze a controversial current debate because it is needed to make a preliminary study to a more in-depth study, or because there are few resources available at the time. The workflow is shown in Figure 4.2.

Comparing the extended Scheme presented in Figure 4.1 with the reduced Scheme presented in Figure 4.2, it can be seen how the reduced version does not analyze the languages used, study the polarity present in the publications and detect the different named entities appearing in the posts. These disclaimers limit the ability to understand the available data but do not affect the ability of the proposed methods to categorize the different posts.

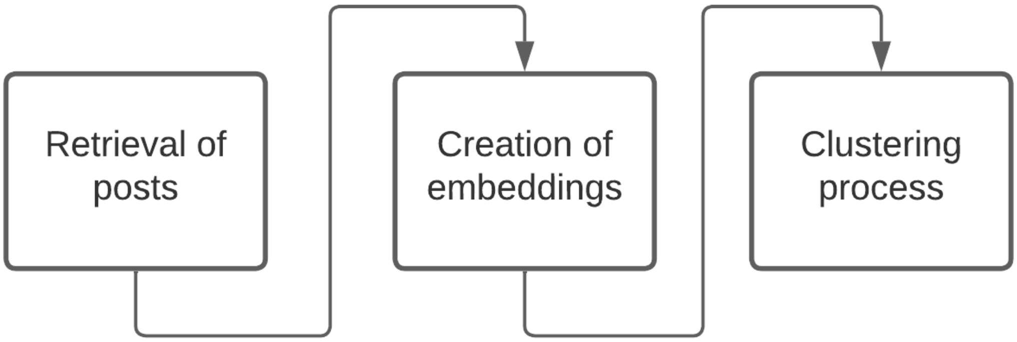


Figure 4.2: Outline of the reduced approach to user profiling.

CHAPTER 5

NLP Analysis

In order to better understand the characteristics of the dataset retrieved for this master's thesis, two natural language processing tasks have been carried out on the captions of the retrieved publications; these tasks are sentiment analysis and named entity recognition.

5.1 Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique that aims to determine the emotional tone or sentiment expressed in a piece of text. To do this analysis, we have used XLM-roBERTa [28] from the Research Group in Natural Language Processing at Cardiff University [49] and fine-tuned for Twitter. The main reasons why we have decided to use this model are:

- Twitter and Instagram are both social networks with the importance of hashtags and user mentions.
- Typically, tweets are short units of text that may contain mentions to other users, URLs, hashtags, and emoticons. Instagram captions share these characteristics. There needs to be more data to create a sentiment analysis model from the retrieved dataset.
- As far as the authors are aware, there is no open-source sentiment analysis model fine-tuned for Instagram

In total, we have managed four different categories:

- Positive (P): The caption shows a positive polarity.
- Neutral (NEU): The captions present no polarity or mixed polarities.
- Negative (N): The caption presents a negative polarity.
- Not applicable (NA): The caption contains no text, showing no polarity.

5.1.1. Post level

For the first task, we conducted a sentiment analysis at the post level; each post was associated with a specific polarity. The main reason for this kind of analysis is that the framework proposed in Section 4 does user profiling work on the post level. Therefore, with this approach, we can study how each kind of user expresses the polarity of their posts. Table 5.1 shows examples of sentences categorized into each polarity.

Polarity	Sentence
P	"Tus palabras serán escuchadas, tu voz será reconocida, tu amor será recibido - #noalaborto #principiosdevida #alzatuvoz #papaymama #provida #diseñodelcielo #libreparamar #hijosconproposito #dejalonacer #conmishijosnotemetas
P	el mejor regalo que pedimos salud vida amor y la bendición de Dios siempre para nuestras familias y las de todo el mundo #11meses ... #mellos #amor #likefor #viralvideos #viralpost #santamarta #gemelosfantasticos #niñoyniña #amordepapas #marineros #respeto #noalaborto #hermanos #mellizos #lunes s #amoamispadres #velitas #gemelar #diciembre #mellospomaresacosta #navidad #papaymama #7dediciembre #picapiedras #vida #salud #amor #doscorazones #like #likeforlikes
NEU	#provida #promujer #proaborto #abortolegalseguroygratuito #abortoseguro #abortolibre #abortolegalya #sororidad #feminismo #feminista #vivasnosqueremos #sialavida #noalaborto #machismo #abortolegalya #proteger #niunamenos #yodecido #patriarcado #amar #abortolegalparanomorir #aborto #abortolegal #vida #cuidar #buentermino #abortolegalesvida #salvar
NEU	San Esteban, Protomártir en cuanto fue el primero en derramar la propia sangre por Cristo. Su nombre significa "coronado". Fue elegido junto con otros 6 diáconos como colaborador de los Apóstoles y murió lapidado. La Iglesia lo celebra el 26 de diciembre. #santamariamadrededios #Dios #Jesus #Cristo #Jesucristo #espiritusanto #virgenmaria #sagradafamilia #iglesiacatolica #fecatolica #catolica_espiritualidad #paz #amor #divinamisericordia #papafrancisco #sanjuanpabloII #sanesteban #navidad #natividad #catecismo #sanfranciscodeasis #sanagustin #batallacultural #provida #noalaborto #instacatolico #catholic #navidad #libros #sanjuanbosco
N	Los Rockefeller financiaron el aborto en la #onu #noalabortosialavida #noalaborto #onucriminal
N	Ah c4br0n... No sabía que enojarse era un privilegio. Buro-Chan. • • • • • #provida #vidas #feminismo #sialavida #aborto #noalaborto #abortolegalya #salvemoslas #niunamenos #abortolegal #salvemoslasdosvidas #feminista #abortoilegal #patriarcado #nadiemenos #conmishijosnotemetas #abortono #olaceleste #machismo #patriarcado #prolife #proaborto #abortolegalparanomorir #abortolegalseguroygratuito #noesno #nofueley #feminazi #feminazis #seraley #padre

Table 5.1: Example of captions and their polarity.

Of the 5865 posts that comprise our dataset, 1164 have a positive polarity, 2052 have a neutral polarity, 2559 have a negative polarity, and 90 posts have no caption; therefore, this analysis is not applicable. The percentages of membership in each category are shown in Figure 5.1.

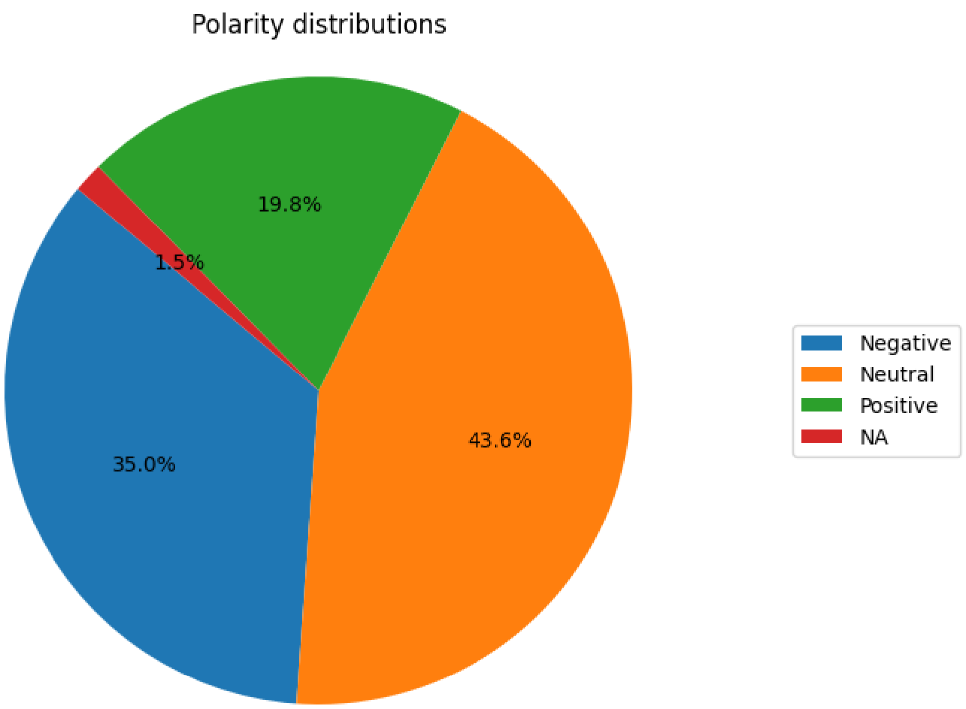


Figure 5.1: Polarity distribution at post level.

The most frequent category in our dataset is "Neutral", with 43.6% of the publications belonging to this category, and "Negative" is the second category, with 35% of the publications. On the other hand, the "Positive" category contains only 19.8% of the publications. Finally, 1.5% of the publications have an empty caption, so this analysis cannot be applied to them. Although the large number of publications with a neutral polarity may be shocking, it is important to remember that this category includes both publications with no polarity and those with mixed polarities, i.e., publications in which both negative and positive polarity are present. This behavior is consistent with that of a debate in social networks, where in the same post, one sentence can be used to defend one's own position and the next to disparage the opposing opinion. With respect to positive and negative polarities, it is noteworthy that there is a greater number of publications with negative polarity than with positive polarity. This fact suggests that in publications where only one polarity is present, captions tend to focus on negative feelings.

The large number of publications classified as neutral has multiple causes:

- The existence of publications whose only text is mentions, emoticons, or hashtags. An example is shown in Table 5.1.
- The existence of publications containing some phrases to attack the right to abortion and other phrases to defend the rights of unborn children.
- Neutral category agglutinates posts with no polarity and with mixed polarity.

5.1.2. Comment level

In order to improve the understanding of the retrieved dataset, we have also performed a sentiment analysis at the comment level. The table 5.2 shows examples of comments that have been classified according to the polarity present.

Polarity	Sentence
P	Feliz Navidad!
P	Yo tengo problemas de irá (que debo tratar) así que soy muy privilegiado
NEU	Qué te pasa? - "nada"
NEU	#provida #noalaborto #salvelas2vidas
N	El descontrol
N	Si, un desastre son enojadas.

Table 5.2: Example of comments and their polarity.

Of the 23208 comments that make up our dataset, 6206 comments have a neutral polarity, 10384 have a negative polarity, 6610 have a positive polarity, and eight comments have a text and therefore cannot be analyzed. The percentages of membership in each category are shown in Figure 5.2.

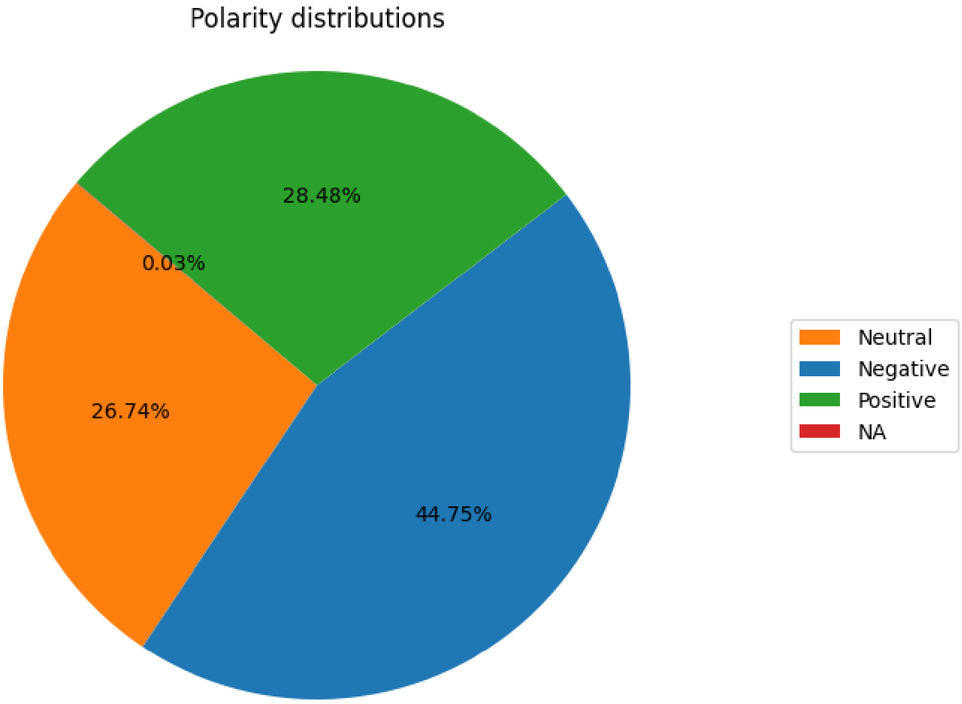


Figure 5.2: Polarity distribution at comment level.

The Neutral and Positive categories have almost the same percentage of publications with 26.74% and 28.48%. Meanwhile, only 0.03% of the comments are not suitable for sentiment analysis. Most of the comments present a negative polarity. one of the reasons for this fact is the existence of publications that attack a position of the debate, and the comments of this type of publication usually defend this attack, thus showing a negative polarity. Regarding the large change in the percentage of neutral publications between posts and comments, it is important to note that comments have a tendency to be shorter than publications, so there is less space for comments that present both positive and negative polarity and are classified as neutral.

5.2 Named entity recognition

In order to perform the named entity recognition task, we have used the natural language processing tool spaCy [50]. Spacy is a library for advanced natural language processing in Python and Cython. spaCy comes with pre-trained pipelines and currently supports tokenization and training for more than 70 languages. It features state-of-the-art speed and neural network models for labeling, parsing, named entity recognition, text classification, and multi-task learning with pre-trained transformers such as BERT. Specifically, we have used the multi-language pipeline "xx_ent_wiki_sm." This pipeline is trained for the named entity recognition task with the WikiNER dataset [51]. This dataset contains 7,200 manually-labeled Wikipedia articles across nine languages: English, German, French, Polish, Italian, Spanish, Dutch, Portuguese, and Russian. The pipeline used classifies the entities present in the text into four categories:

In total, we have managed four different categories:

- Persons.
- Organizations.
- Miscellany.
- Locations.

5.2.1. Post level

Of the 5865 posts that comprise our dataset, we have detected 7377 unique named entities that, in total, appear on 40208 occasions. Some named entities fall into several categories depending on the context in which they appear. For example, "San José'" can belong to both the category persons and the category place. Table 5.3 compiles the number of entities retrieved from each category and the number of appearances. Table 5.4 shows the top 10 entities of each category that appear more frequently.

Category	Unique named entity	Number of appearances
Persons	1514	5256
Organizations	1067	3445
Miscellany	3611	20315
Locations	1435	11192

Table 5.3: Named entities categories: number and frequency

A greater number of unique named entities in a category does not imply a greater number of occurrences of named entities in that category. In order to facilitate the data analysis, Figure 5.3 presents the percentage of occurrence of each category present in the named entities.

The category with the highest number of occurrences is Miscellany, which groups all entities that do not belong to the other categories. The second most frequent category is Locations, which brings together all the detected locations. These locations are not limited only to countries or continents; they can refer to local areas, neighborhoods, etc. In the debate on abortion, the origin of the messages is important since, depending on the country or even region of the country, the legal status of abortion can vary greatly. It is also noteworthy that more references are made to individuals than to organizations. This fact is striking since in the 3.4.3 section, it has been observed that several of the most active accounts are related to religious organizations.

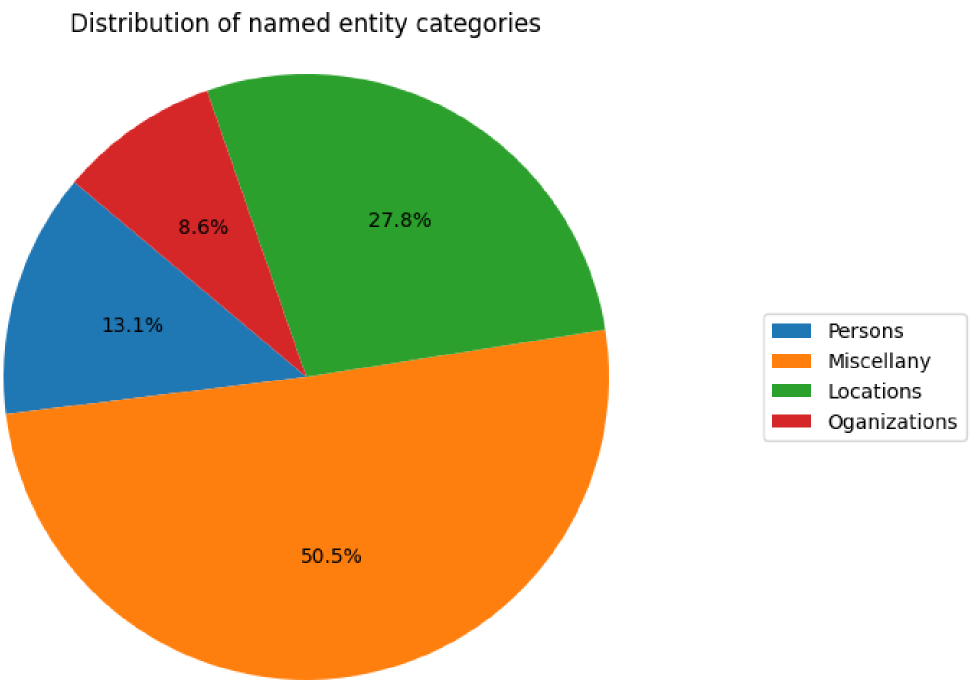


Figure 5.3: Named entity categories distribution.

Persons	Organizations	Miscellany	Locations
Cristo	NoAlAborto	QuintanaRoo	Venezuela
Jesús	FueraCastillo	NoalAborto	Colombia
Jesus	RafaelLopezAliaga	SiALaVidaQRoo	Argentina
Jesucristo	Iglesia	BajaCaliforniaSur	Brasil
Noalaborto	VIDA	Follow	Mexico
jesus	NadieMenos	Dios	Nicaragua
Señor	SI	nofueley	Perú
María	ProyectoAngel	RenovacionPopular	Peru
Padre	ABORTO	NoalComunismo	VacanciaPedroCastillo
Abortonuncamas	IglesiaCatolica	NuevoOrdenMundial #Noalnuevoordenmundial	mexico

Table 5.4: Top 10 most frequent entities by category in posts.

As noted in Section 3.4.2, a clear relationship exists between the Christian religion and anti-choice profiles and publications. In the named entities recognition, this fact stands out, especially in the category of people, where eight of the ten most frequent entities are directly related to biblical figures or classical ways of referring to god. Within the category of organizations, this frequency is also noticeable due to the large number of appearances of entities such as Iglesia or IglesiaCatolica. In the miscellaneous category, this fact is also observed with the appearance of the entity "Dios" (God).

Regarding the miscellaneous category, the detection of "Follow" stands out, which refers to the hashtag "Follow," a hashtag used to increase the profile's reach in the social network.

Concerning the Locations category, it is worth noting that the most repeated entities refer to countries.

5.2.2. Comment level

Of the 23208 comments that comprise our dataset, we have detected 7464 unique named entities that, in total, appear on 16924 occasions. Some named entities fall into several categories depending on the context in which they appear. For example, "San José'" can belong to both the category persons and the category place. Table 5.5 compiles the number of entities retrieved from each category and the number of appearances. Table 5.6 shows the top 10 entities of each category that appear more frequently.

Category	Unique named entity	Number of appearances
Persons	1803	3860
Organizations	1019	1982
Miscellany	3666	8305
Locations	1255	2777

Table 5.5: Named entities categories: number and frequency

Unlike what happened when we carried out the analysis at the publication level in the comments, a greater number of unique named entities in a category is linked to a greater number of appearances in said category. This is because, in the comments, the named entities have less tendency to repeat themselves. In order to facilitate the data analysis, Figure 5.4 presents the percentage of occurrence of each category present in the named entities.

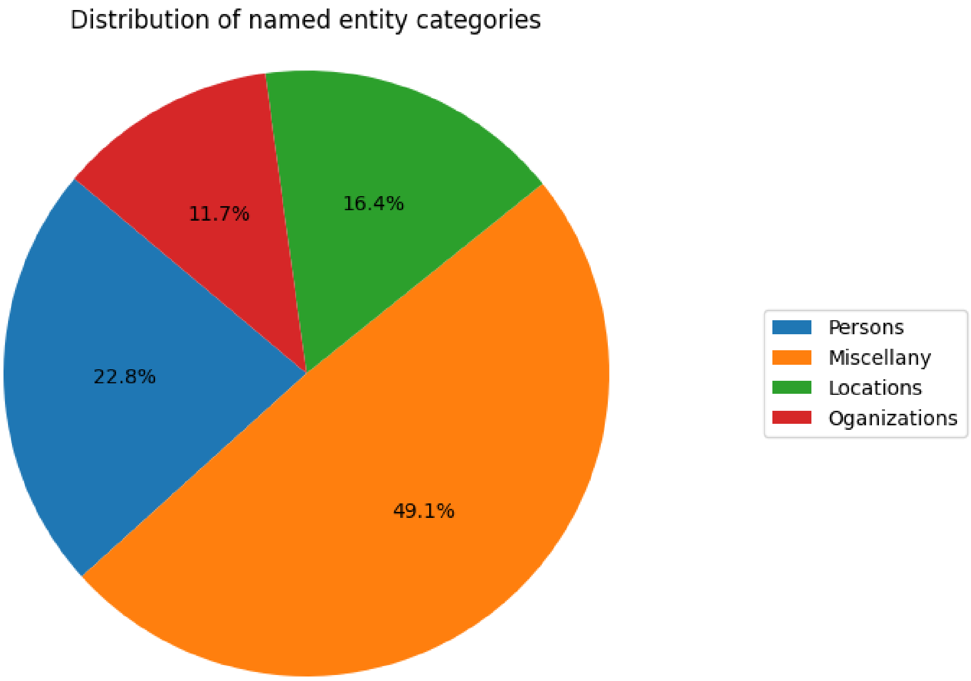


Figure 5.4: Named entity categories distribution.

The category with the highest number of occurrences is Miscellany, which groups all entities that do not belong to the other categories. The second most frequent category is Locations, which brings together all the detected locations. These locations are not limited only to countries or continents; they can refer to local areas, neighborhoods, etc. In the debate on abortion, the origin of the messages is important since, depending on the country or even region of the country, the legal status of abortion can vary greatly.

It is also noteworthy that more references are made to individuals than to organizations. This fact is striking since in the 3.4.3 section, it has been observed that several of the most active accounts are related to religious organizations.

Persons	Organizations	Miscellany	Locations
Jesús	broken heart emoji	Dios	Argentina
Lucio	person facepalming emoji	small white heart emoji	Chile
Cristo	DIOS	3x small white heart emoji	Amén
Amén	NO	green heart emoji	Precio
Señor	OK hand emoji	square emoji	face with open mouth emoji
Bueno	ABORTO	2x small white heart emoji	chile
Send	OMS	@youngesthuman	Provida
woman shrugging emoji	ONU	@shopipi_3062	NO
Asco	NoAlAborto	man facepalming emoji	Amen
Jajajaja	sparks emoji	woman facepalming emoji	face vomiting

Table 5.6: Top 10 most frequent entities by category in comments.

As noted in Section 3.4.2 and in Section 5.2.1, a clear relationship exists between the Christian religion and anti-choice profiles and publications. In the named entities recognition, both at the post level and at the comment level, this fact stands out, especially in the category of people. In the comments, the appearance of entities directly related to Christianity is less than in the publications, although it is still notable. The most interesting characteristic revealed during this analysis is the increase in the use of emoticons with respect to posts. These emoticons appear in multiple categories, and it can be seen that they are generally emoticons with strong polarity charges.

CHAPTER 6 Embeddings

Embeddings are continuous vector representations of real-world objects and relationships. They aim to convert complex and categorical data into numerical vectors while preserving the essential characteristics of the original data. High-quality embeddings enable machines to operate effectively with symbolic and categorical information. Instagram posts are composed of both textual information and information in the form of images and videos.

Instagram posts may contain text, photographs, or video, so means of encoding all of these types of information are necessary. However, the amount and quality of information extractable from each type of content is different. For this reason, we propose three ways to generate the embeddings that represent the available information:

- Textual embeddings: In this approach, we use only the text from the captions of the publications.
- Image and video embeddings: In this approach, we use only the media content from publications.
- Multimodal representations.: In this approach, we use both textual and media information to represent the publications.

Once we have generated the three types of embeddings, we will proceed to evaluate in Section 8 which ones are the most appropriate for analyzing the abortion debate on Instagram.

6.1 Text embeddings

How to represent textual information to feed machine learning and deep learning algorithms has been a widely studied field. Generally, two types of embeddings are distinguished: contextual embeddings and non-contextual embeddings.

Non-contextual embeddings represent each token always with the same vector, independently of its context [10]. By not using contextual information for the creation of embeddings, any set of words that appear together is represented independently of the context. For example, the set of words "I want to play" would be encoded in the same way in the sentences "I want to play basketball" and "I want to play the guitar". This approach is based on the assumption that words that occur in similar contexts tend to have a similar meaning so that the concept of context is directly integrated into the representation created [10].

The most known approach for non-contextual embeddings is the one-hot representation, where each component corresponds to each term in a vocabulary. Therefore, a sparse vector is obtained that has a size equal to the size of the vocabulary, and each constituent of the vector represents a concept. Another example of non-contextual embeddings is the Continuous Bag-of-Words Model (CBOW). It is an architecture that evolves from the feedforward NNLM [52]. CBOW consists of input, projection, hidden, and output layers. At the input layer, N previous and M posterior words are encoded using 1-of-V coding, where V is the size of the vocabulary. The input layer is then projected to a projection layer P that is shared across all the words, and finally, all words get projected into the same position (their vectors are averaged) [52].

Contextual embeddings represent each word by different embeddings depending on the context of the word, e.g., the set of words "I want to play" would be encoded differently in the sentences "I want to play basketball" and "I want to play the guitar". This approach can naturally model complex features of the tokens depending on specific contexts such as polysemy, coreference, etc. The approaches intended to compute contextual representations derive them from the hidden layers of some kind of neural encoder applied on sequences of tokens, and they are pretrained and then fine-tuned on downstream tasks [10].

BERT [22] is a Transformer encoder trained on two self-supervised objectives: masked language model and next sentence prediction. It was the first work intended to pretrain a Transformer model on large corpora by means of two pretraining objectives: Masked Language Model (MLM) and Next Sentence Prediction (NSP). On the one hand, MLM is basically a cloze task where random tokens are masked, forcing the model to use the bidirectional context of a given masked token to predict it. This objective, along with the Transformer encoder, allows BERT to naturally model bidirectional contextual representations. On the other hand, the NSP signal was proposed with the aim of learning the coherence by means of a binary classification, which consists in determining if a text segment A precedes a text segment B in the source [10]. The research carried out around BERT can be condensed into two different avenues of research: the general improvement of the model and the fine-tuning of the model for tasks other than the original ones. Regarding the first path, the most important one are the AlBERT model [53], which employs a SOP signal to improve the inter-sentence coherence of the model.SOP is a reformulation of NSP where pairs of unordered sentences are used to force the model to learn inter-sentence coherence instead of topic coherence as induced by NSP. Regarding the MLM objective, SpanBERT [54] proposed several span masking strategies, using a span boundary objective for predicting each token in a masked span using the tokens on its boundary. Finally, RoBERTa [24], is a BERT model with a careful design of its hyperparameters, training corpora, and practical strategies. The main novelties of RoBERTa were not considering the NSP signal, a dynamic masking strategy, instead of defining a single masking pattern for each sample, and training with large batches, which was shown to improve the perplexity in the MLM objective as well as the performance on downstream tasks [10].

E5 [55], EmbEddings from bidirEctional Encoder rEpresentations, is a family of state-of-the-art text embeddings that transfer well to a wide range of tasks. E5 is a family of models created especially for the creation of textual embeddings. This family is formed by three models that differ from each other by the number of final parameters and the hyperparameters used. Specifically, the E5_{small} model has 33M parameters, E5_{base} 110M and E5_{large} 330M. The training was conducted in two phases. In the first phase, the model was pre-trained by means of contrastive pretraining using CCPairs [55], a curated web-scale text pair dataset containing heterogeneous training signals, to achieve this goal. The CCPairs dataset is constructed by combining various semistructured data sources such as

CommunityQA, Common Crawl, and Scientific papers and performing aggressive filtering with a consistency-based filter to improve data quality. This method has been used so that the model learns to distinguish pairs of relevant texts to improve performance from those that have no impact or even a negative impact [**?**]. In the second phase, the model was fine-tuned using supervised learning using a combination of three different datasets: NLI (Natural Language Inference), MS-MARCO passage ranking dataset [56], and NQ (Natural Questions) dataset [57, 58]. The authors of [55] evaluated the model using extensive experiments on both BEIR and MTEB benchmarks, demonstrating the method's effectiveness. On the BEIR zero-shot retrieval benchmark [59], E5 is the first model to outperform the strong BM25 baseline without using any labeled data. When fine-tuned on labeled datasets, the performance can be further improved. Results on 56 datasets from the recently introduced MTEB benchmark [60] show that our E5base is competitive against GTRxxl and Sentence-T5xxl, which have 40× more parameters. All these experiments demonstrate that E5 models can be readily used as a general-purpose embedding model for any tasks requiring a single-vector representation of texts, such as retrieval, clustering, and classification, achieving strong performance in both zero-shot and fine-tuned settings.

The model chosen to generate the embeddings is Multilingual-E5-small because:

- Multilingual model: The model is multilingual, and in our dataset, we have posts with multiple languages, as seen in Section 3.4.4.
- The model has been tested for zero-shot tasks, and we do not have labels for fine-tuning.
- The model has been trained as a general-purpose embedding model.

6.2 Image and video embeddings

In the second approach, we will only use the media (images and videos) present in the posts to represent the publications.

One of the most common approaches to creating visual embeddings has been the use of region-based features produced by an out of the box object detection network [61]. Object detection is an important computer vision task that deals with detecting instances of visual objects of a certain class (such as humans, animals, or cars) in digital images [62]. This ability to sort parts of a larger image is especially useful when creating visual embeddings. It allows the overall embedding of an image to contain information about the different objects that appear in the image.

There are mainly two groups of deep learning techniques for this task [62]:

- One-stage detector: This category includes techniques that capture all the elements present in the image in a single step. Generally, techniques in this category sacrifice performance in exchange for an extremely short inference time.
- Two-stage detector: This category groups together techniques that perform a refinement process on the detected objects. They are generally techniques that achieve high precision in exchange for high inference times.

YOLO [63] is one of the main models of the one-stage detector approach. Their authors have reframed the problem of object detection as a regression problem instead of a classification problem. A convolutional neural network predicts the bounding boxes as well as class probabilities for all the objects depicted in an image. As this algorithm identifies the objects and their positioning with the help of bounding boxes by looking at the image only once, hence they have named it as You Only Look Once (YOLO) [64]. In order to detect the different objects in an image, the model divides the image into grid regions and proceeds to predict the probability of each class in each of the cells of the grid. So, for each cell of the image, the probability that each class is within that cell is processed. Adjacent grid cells may also predict the same object, i.e., predicting the overlapping bounding boxes for the same object. So, there would be multiple predictions because neighboring grid cells may assume the object center falls inside it. After that, all predictions with a probability under a threshold are discarded. After discarding, the bounding boxes of each overlapping class are unified. The second criteria for discarding the less relevant bounding boxes is known as non max suppression, which is further based upon the Intersection over UNion (IoU). First, the box with the maximum class score is selected. All other bounding boxes overlapped with the chosen box will be discarded having IoU is greater than some predefined threshold.. These steps are repeated until there are no bounding boxes with lower confidence scores than the chosen bounding box. The main advantage of the YOLO model is the high inference speed; on the other hand, YOLO suffers from a drop in localization accuracy compared to two-stage detectors, especially for some small objects [62].

In 2015, Girshick [65] proposed a two-stage Fast RCNN detector. This approach is based on convolutional neural networks (CNN). In order to perform the object detection, the model processes the whole image using several CNN layers to produce a convolutional feature map. In the second stage, for each object proposal, a region of interest (RoI) pooling layer extracts a fixed-length feature vector from the feature map. Then, each feature vector is fed into a sequence of fully connected layers that finally branch into two sibling output layers: one that produces softmax probability estimates over K object classes plus a catch-all "background" class and another layer that outputs four real-valued numbers for each of the K object classes [62].

In addition to this approach, a new one has emerged due to the influence of the successes obtained by transformer models in NLP tasks. Transformers discard the traditional convolution operator in favor of attention-alone calculation in order to overcome the limitations of CNNs and obtain a global-scale receptive field [62]. The Vision Transformer (ViT) presented in [39] is pure a transformer encoder model (BERT-like) pre-trained on an extensive collection of images in a supervised fashion, namely ImageNet-21k, at a resolution of 224x224 pixels. Next, the model was fine-tuned on ImageNet (also referred to as ILSVRC2012), a dataset comprising 1 million images and 1000 classes, also at resolution 224x224. Images are presented to the model as a sequence of fixed-size patches (resolution 16x16), which are linearly embedded. By pre-training the model, it learns an inner representation of images that can be used to extract features useful for multiple tasks such as image classification or clustering [39].

While transformers designed for NLP tasks use text sequences as input, transformers for computer vision tasks receive images with an arbitrary number of channels as input. In order to handle the images, vision transformers segment the input into a flattened sequence of 2D patches whose size depends on the number of channels in the image and the resolution of each patch and the original image. Following the approach designed for BERT [22], a token *[class]* is used, which is embedded in the image and is learnable. The state of this embedding serves as a representation of the image. When it is desired to train the model or apply fine-tuning, classifier heads are added along with one-dimensional embeddings to maintain positional information. Like transformer models designed for NLP tasks, ViT is usually pre-trained on large datasets and then fine-tuned for each task using smaller datasets [38].

ViT yields modest results when trained on mid-sized datasets such as ImageNet, achieving accuracies of a few percentage points below ResNets of comparable size. Because transformers lack some inductive biases inherent to CNNs such as translation equivariance and locality–they do not generalize well when trained on insufficient amounts of data. However, the authors found that training the models on large datasets (14 million to 300 million images) surpassed inductive bias. When pre-trained at a sufficient scale, transformers achieve excellent results on tasks with fewer data points. For example, when pre-trained on the JFT-300M dataset, ViT approached or even exceeded state of the art performance on multiple image recognition benchmarks. Specifically, it reached an accuracy of 88.36% on ImageNet and 77.16% on the VTAB suite of 19 tasks.

The transform models popularized for natural language processing tasks have also had a major impact on video-based computer vision tasks. When working with videos, there are two main problems that need to be addressed:

- Temporal redundancy: In a video, captured frames are frequent. Semantics vary slowly in the temporal dimension [66]; consecutive frames are highly redundant.

- Temporal correlation: Videos can be seen as the temporal extension of the static appearance. Therefore, there is an inherent correspondence between adjacent frames [42].

In [42], VideoMAE is proposed. VideoMAE is an extension of Masked Autoencoders (MAE) to video. The model's architecture is very similar to a standard Vision Transformer (ViT), with a decoder on top for predicting pixel values for masked patches. Videos are presented to the model as a sequence of fixed-size patches (resolution 16x16), which are linearly embedded. By pre-training the model, it learns an inner representation of videos that can be used to extract features useful for multiple tasks such as video classification or clustering.

In more detail, VideoMAE proposes a processing based on four steps [42]:

- Temporal downsampling: due to temporal redundance intrinsic to the videos, the authors of [42] propose to use a strided temporal strategy to sample the data.

- Cube embedding: In order to reduce the impact of temporal correlation, joint space-time cube embedding is used so that the data is not time-stripped.

- Tube masking with extremely high ratios: Tube masking to an extremely high ratio is proposed with the objective of reducing the impact of temporal correlation. Temporal tube masking consists in making the masks distributed on the temporal axis, in addition to the temporal correlation.

- Backbone: joint space-time attention. Due to the high proportion of masking ratio, only a few tokens are left as the input for the encoder. To better capture high-level spatiotemporal information in the remaining tokens, the authors propose the use of a vanilla ViT backbone and adopt joint space-time attention. Thus, all pair tokens could interact with each other in the multi-head self-attention layer. This backbone is first pre-trained with image data in a supervised form. Then, these backbones are fine-tuned for downstream tasks [42].

The models chosen to generate the embeddings are Vision Transformer (ViT) [39] for images and Video Masked Autoencoder (VideoMAE) [42] for video.

6.3 Multimodal embeddings

In the third approach, we will create embeddings from a multimodal perspective, using both the images and the captions present in the posts to create the embeddings.

Vision language modeling (VL) is the domain where computer vision and natural language processing intersect [61]. Visual language modeling is a key element to be able to develop tasks such as visual question answering, image captioning, or multimodel embeddings. As in many research fields, transformers have improved the results obtained on the previous paradigms by pretraining models on large datasets of image-text pairs before transferring them to other tasks, usually with minor changes to parameter values and architecture.

Mainly, four variations of the architecture have been used in the literature [61]:

- Dual encoders: Dual encoders model visual and textual representations separately, and the modalities do not interact within the deep learning model. Instead, the output of the visual and textual modules interact through a simple mechanism, usually a cosine similarity [61]. One example of a model following this architecture is ALIGN [43], which will be further explained later in this section.
- Fusion encoders: There are two approaches within the family of fusion encoders [61]:
  - Single-tower architecture: Single transformer encoder operates on a concatenation of visual and textual input representations. Since both the visual and textual tokens are embedded into a single input, the single transformer stack allows for unconstrained modality interaction modeling. One example of a model following this architecture is VL-BERT [67].
  - Two-tower architecture: Each modality is in separate transformer stacks, and interaction is then achieved through a cross-attention mechanism. One example of a model following this architecture is LXMERT [68].
- Combination encoders: These models contain separate visual and textual encoders at the base of the model. The outputs of the text encoder and an image encoder are aligned using cosine similarity before being fed into a fusion encoder module of some kind. One example of a model following this architecture is FLAVA [69].
- Encoder decoder models: Following the architecture of the original transformer, some VL models opt for a design consisting of at least one encoder stack and a decoder stack. This model architecture is versatile in general and allows models using them to successfully perform a wide range of functions, including generative tasks such as image captioning. One example of a model following this architecture is OmniVL [70].

ALIGN is the model chosen to generate the embeddings, presented in [43]. ALIGN model consists of a pair of image and text encoders with a cosine-similarity combination function at the top. Specifically, EfficientNet with global pooling (without training the 1x1 convolutional layer in the classification head) as the image encoder and BERT with *[CLS]* token embedding as the text embedding encoder. A fully connected layer with linear activation is added on top of BERT encoder to match the dimension from the image tower. Image and text encoders are learned via a contrastive loss (formulated as normalized softmax) that pushes the embeddings of matched image-text pairs together while pushing those of non-matched image-text pairs apart. The visual representations created by this model achieve strong performance when transferred to classification tasks such as ImageNet and VTAB. The aligned visual and language representations enable zero-shot image classification and set new state-of-the-art results on Flickr30K and MSCOCO image-text retrieval benchmarks, even when compared with more sophisticated crossattention models.

CHAPTER 7
Clustering

Clustering is an unsupervised machine learning technique that involves grouping similar data points into clusters based on specific inherent patterns or similarities. Clustering plays an essential role in detecting, organizing, and understanding what types of data are present in our dataset. For the current task, it helps to categorize the types of posts that are being published on the debate on the right to abortion. From the number of clusters perspective, there are two kinds of techniques:

- Predefined number of clusters techniques: Some techniques need to know how many clusters they should divide the dataset. These techniques are helpful when there is a broad prior knowledge of the task and the many types of data in the available dataset. Some examples of these techniques are K-Means, Birch, and Spectral-Clustering.
- Automatic discovery of clusters techniques: some techniques do not need to know how many clusters exist on a dataset. These techniques are helpful when there is no prior knowledge of how many categories there are present in the data. Some examples of this kind of technique are HDBSCAN, OPTICS, and DBSCAN

As we do not have any prior knowledge about how many types of posts there are, we need to use techniques from the second family presented above.

7.1 Dimensionality reduction

When we want to deal with highly changing situations in which one of the priorities is to obtain quick high quality results, using techniques that use small representation spaces is especially useful, since techniques generally work faster when using small representation spaces. So we are going to apply dimensionality reduction techniques on the embeddins obtained using the techniques presented in 6 to study the quality of results when using small representation spaces. To reduce the dimensionality of the data points, we are using UMAP (Uniform manifold approximation and projection).

UMAP is a general-purpose manifold learning and dimension reduction algorithm. It provides a general framework for approaching manifold learning and dimension reduction but can also provide specific concrete realizations. UMAP identifies a pre-set number of nearest neighbors and represents distances to these neighbors as a weighted graph where the nearest neighbors are weighted more heavily. The goal is to find a low-dimensional representation of the data that preserves these neighborhoods as much as possible. By focusing on preserving neighborhood topology rather than absolute distances, UMAP allows for data-dense regions to be “stretched out” in the representation.

This can have the benefit of reducing overcrowding of the low-dimensional representation but comes at the cost of a more challenging interpretation of distances [71].

7.2 Clustering techniques

In order to create the best clusters to categorize the post, we have used and evaluated three techniques: HDBSCAN, DBSCAN, and OPTICS.

The DBSCAN algorithm [72] views clusters as areas of high density separated by low-density areas. Due to this rather generic view, clusters found by DBSCAN can be any shape, as opposed to other cluster methods such as k-means, which assumes that clusters are convex-shaped. The central component of the DBSCAN is the concept of core samples, which are samples in high-density areas. A cluster is, therefore, a set of core samples, each close to each other (measured by some distance measure we are using) and a set of non-core samples that are close to a core sample (but are not themselves core samples).

The OPTICS algorithm [73]shares many similarities with the DBSCAN algorithm and can be considered a generalization of DBSCAN that relaxes the distance requirement. IN DBSCAN, the maximum distance between two samples for one to be considered as in the neighborhood of the other is a single value; meanwhile, in OPTICS, it is a value range. The key difference between DBSCAN and OPTICS is that the OPTICS algorithm builds a reachability graph, which assigns each sample both a reachability_distance and a spot within the cluster. The reachability distances generated by OPTICS allow for variable density extraction of clusters within a single data set.

The HDBSCAN algorithm [74, 75] can be seen as an extension of DBSCAN and OPTICS. DBSCAN assumes that the clustering criterion (i.e., density requirement) is globally homogeneous. In other words, DBSCAN may struggle to capture clusters with different densities successfully. HDBSCAN alleviates this assumption and explores all possible density scales by building an alternative representation of the clustering problem.

7.3 Metrics

In order to evaluate the different cluster methods and the different ways to generate embeddings, three metrics are going to be used: the Silhouette Coefficient, the Davies-Bouldin index, and the Calinski-Harabasz Index.

7.3.1. Silhouette Coefficient

Silhouette Coefficient [76] evaluates clusters by function of their tightness and separation. This silhouette shows which objects lie well within their cluster and which ones are merely somewhere in between clusters. The mathematical formulation for one sample can be seen at 7.1. The Silhouette Coefficient for a set of samples is given as the mean of the Silhouette Coefficient for each sample.

$$\text{Silhouette Coefficient} = \frac{b-a}{\max(a,b)} \tag{7.1}$$

Where a is the mean distance between a sample and all other points in the same class, and b is the mean distance between a sample and all other points in the next nearest cluster.

The main advantages of this metric are:

1. The metric is easily interpretable as the score is bounded between -1 for incorrect clustering and +1 for highly dense clustering. Scores around zero indicate overlapping clusters.

2. The score is higher when clusters are dense and well separated, which relates to a standard concept of a cluster

The main disadvantages are:

1. The Silhouette Coefficient is generally higher for convex clusters than other concepts of clusters, such as density-based clusters like those obtained through DBSCAN.

7.3.2. Davies-Bouldin index

Davies-Bouldin index [77] signifies the average 'similarity' between clusters, where the similarity is a measure that compares the distance between clusters with the size of the clusters themselves. The index is defined as the average similarity between each cluster and its most similar context. This similarity is defined according at Equation 7.2 and then Davies-Bouldin score is defined according to Equation 7.3.

$$R_{ij} = \frac{s_i + s_j}{d_{ij}} \tag{7.2}$$

Where s_i is the average distance between each cluster point i and the centroid of that cluster, also known as cluster diameter, and d_ij is the distance between clusters centroids i and j.

$$DB = \frac{1}{k} \sum_{i=1}^{k} \max_{i \neq j} R_{ij} \tag{7.3}$$

The main advantages of this metric are:

1. The metric is easily interpretable as the score as values close to zero indicates better partitions.

2. The index is solely based on quantities and features inherent to the dataset as its computation only uses point-wise distances.

The main disadvantages are:

1. The Davies-Boulding index is generally higher for convex clusters than other concepts of clusters, such as density-based clusters like those obtained from DBSCAN.

2. The usage of centroid distance limits the distance metric to Euclidean space.

7.3.3. Calinski-Harabasz index

Calinski-Harabasz index [78] is the ratio of the sum of between-clusters dispersion and of within-cluster dispersion for all clusters (where dispersion is defined as the sum of distances squared). The mathematical formulation can be seen at 7.6.

$$s = \frac{\mathrm{tr}(B_k)}{\mathrm{tr}(W_k)} \times \frac{n_E - k}{k - 1} \tag{7.4}$$

$$W_k = \sum_{q=1}^{k} \sum_{x \in C_q} (x - c_q)(x - c_q)^T \tag{7.5}$$

$$B_k = \sum_{q=1}^{k} n_q (c_q - c_E)(c_q - c_E)^T \tag{7.6}$$

Where $\mathrm{tr}(B_k)$ is trace of the between group dispersion matrix and $\mathrm{tr}(W_k)$ is the trace of the within-cluster dispersion matrix, C_q is ther set of the points in cluster q, c_E the center of E and n_q the number of clusters in q.-

It is the mean distance between a sample and all other points in the same class, and b is the mean distance between a sample and all other points in the next nearest cluster.

The main advantages of this metric are:

1. The metric is easily interpretable as high values relate to better-defined clusters.
2. The score is higher when clusters are dense and well separated, which relates to a standard cluster concept.
3. The score is fast to compute.

The main disadvantages are:

1. The Calinski-Harabasz index is generally higher for convex clusters than other concepts of clusters, such as density-based clusters like those obtained through DB-SCAN.

CHAPTER 8
Experimentation

We will now present the experimentation carried out and the results obtained. The experimentation has a fourfold objective:

- To analyze which type of representation allows a higher quality non-supervised clustering.
- To analyze which unsupervised clustering algorithm allows to obtain higher quality clusters, keeping the total number of clusters low enough to allow human analysis.
- To study if the effect of dimensionality reduction techniques in the clustering process is significant.
- To study the effect of changing hyperparameters in the clustering process and the discovery of the best hyperparameters.

Currently, we do not have a set of classes in which to categorize each publication in our dataset. Since one of the ongoing work involves a more in-depth study of the abortion debate on Instagram together with experts, it is considered that conducting an unsupervised study may facilitate this work by helping to discover the internal patterns present in the data. Instagram posts contain at least one image and/or video and one caption. Therefore, we have both textual and visual information. This allows us to use the different types of embeddings presented in Section 6; namely, we can use textual embeddings from media content and visual embeddings. Since each type of information can have a different internal structure, we have used three different clustering algorithms so that we avoid the biases present in each algorithm. These three algorithms are DBSCAN, OPTICS, and HDBSCAN, previously explained in Section 7. Along with the explanation of these algorithms, we have also presented UMAP, a dimensionality reduction technique that will allow us to study whether it is beneficial to greatly reduce the representation space, thus accelerating the training and prediction processes of neural networks.

As each clustering algorithm can perform differently depending on the hyperparameters, we have explored the following three:

- Number of neighbors: This parameter determines how many neighbors the dimensionality reduction algorithm looks at to classify each piece of data. It specifically balances the local and global influence of the data in the clustering process. Low values of this parameter force UMAP to focus on the local structure of the data, while high values force UMAP to focus on the global structure. The tested values are: 5, 50, 100.

- Number of components: This parameter allows us to determine the final dimensionality obtained after applying UMAP, the clustering algorithm. The tested values are: 3, 5, 20, 50.
- Minimum samples: The minimum number of data points that must be within a certain distance of a core point for those data points to be considered part of the same cluster. The tested values are: 50, 100, 500, 1000.

We have experimented using all the combinations of embeddings, algorithms, and hyperparameters, giving rise to a total of 540 results, which are shown in their entirety in Annex A. In this Section, we are going to proceed to analyze the results, firstly at the level of embeddings and secondly in a general way.

The following sections will be structured as follows:

1. Reference to full results.
2. Analysis of the results of each clustering algorithm used based on each of the metrics.
3. Analysis of the best results obtained with each clustering algorithm based on each metric.

Once we have analyzed the results based on each type of embedding, we will proceed to analyze the best results obtained with any type of embedding.

8.1 Text embeddings

First of all, we are going to analyze the results obtained using embeddings generated from the captions. The model chosen to create these embeddings is the Multilingual-E5-small model previously presented in Section 6. This model generates embeddings of size 384. All the results obtained using this type of embedding, depending on the clustering algorithm used, can be seen in the following:

- DBSCAN: Results can be found in Table A.1.
- OPTICS: Results can be found in Table A.2.
- HDBSCAN: Results can be found in Table A.3.

8.1.1. DBSCAN

In order to facilitate the presentation and analysis of the results, we have extracted the best results obtained using the DBSCAN clustering algorithm in Table 8.1. For this, we have defined the best results as those configurations that generate the two best results for each metric.

Table 8.1 shows several interesting results. The first one is that the best results for the Davis-Bouldin score metric are achieved with two different configurations, which have two parameters in common. Both use 50 as the number of neighbors and 500 as the minimum samples. The variation is in the number of components, as the best results can be obtained with both 3 and 50 components; when only three components are used, the best results are also obtained for the Calinski-Harabasz metric. On the contrary, for the Silhouette score, the best results are obtained with a completely different configuration,

Neighbours	Components	Min samples	Davies-Bouldin Score	Calinski-Harabasz Score	Silhouette Score	Number of clusters
50	3	500	**1.1071**	**736.7043**	0.1140	2
50	50	500	1.1071	719.8654	0.1092	2
50	5	500	1.1073	732.4475	0.1128	2
100	5	50	2.4574	156.4663	**0.2097**	18
100	50	50	2.4722	155.9141	0.2095	18

Table 8.1: Top results using text embeddings and DBSCAN

where the minimum samples go down to 50, and the number of neighbors goes up to 100. In these results, we see how a good value for the Davies-Bouldin, i.e., a value close to zero, also generates good results for the Calinski-Harabasz metric (in which higher values imply better clusters). The best results with these metrics generate only two clusters, but the configurations that generate the best results for the silhouette score generate 18 clusters, which may be too many to perform an in-depth qualitative analysis.

8.1.2. OPTICS

In order to facilitate the presentation and analysis of the results, we have extracted the best results obtained using the OPTICS clustering algorithm in Table 8.2. For this, we have defined the best results as those configurations that generate the two best results for each metric.

Neighbours	Components	Min samples	Davies-Bouldin Score	Calinski-Harabasz Score	Silhouette Score	Number of clusters
100	3	500	**1.1845**	**730.7169**	0.1139	2
100	20	500	**1.1845**	**730.7169**	0.1139	2
100	50	500	**1.1845**	**730.7169**	0.1139	2
50	50	50	2.2060	149.4639	**0.2189**	20
50	384	50	2.2455	148.3657	0.2158	20

Table 8.2: Top results using text embeddings and OPTICS.

In Table 8.2, we analyze the results obtained with OPTICS here as before; the best results for the one-score baseball metric and for the kalinski metric are obtained using the same configuration. Specifically, three different configurations generate the same results; these three configurations share the number of neighbors and the minimum samples in addition to the name of clusters created. The neighbor's number is 100, 500 is the minimum number of samples, and two different clusters are generated. When we look at the silhouette metric, the best configurations change radically. They use 50 as the number of neighbors and 50 as the size of the cluster number, and 20 different clusters are generated.

8.1.3. HDBSCAN

In order to facilitate the presentation and analysis of the results, we have extracted the best results obtained using the HDBSCAN clustering algorithm in Table 8.3. For this, we have defined the best results as those configurations that generate the two best results for each metric.

Neighbours	Components	Min samples	Davies-Bouldin Score	Calinski-Harabasz Score	Silhouette Score	Number of clusters
100	20	500	**1.185**	**730.717**	0.114	2
100	5	500	1.186	730.088	0.114	2
100	50	500	1.187	730.325	0.114	2
50	50	50	2.421	168.964	**0.218**	17
50	384	50	2.450	167.801	0.215	17
100	3	50	2.363	167.172	0.215	17

Table 8.3: Top results using text embeddings and HDBSCAN.

Table 8.3 presents the results obtained using the HDBSCAN algorithm. Once again, the Davis-Bouldin metric and the Calinski-Harabasz metric share configurations that are better results. It is remarkable that there are three configurations that generate extremely similar quality metrics sharing as parameters the number of neighbors and the minimum size of clusters; the only variable that changes its value is the number of components. Using the third metric, the Silhouette score, the ideal configurations again vary greatly as the minimum samples decrease from 500 to 50. However, as for the Davis-Bouldin and Calinski-Harabasz metrics, the number of components does not seem to have a great impact.

8.1.4. Text embeddings results

Paying attention to the conclusions drawn in the last three subsections, we can observe the following facts:

- When using embeddings coming from text, the Davis-Bouldin and Calinski-Harabasz metrics obtain the best results using similar configurations with minimum samples equal to 500 and 50 or 100 as the number of neighbors. Moreover, these configurations have always generated only two different clusters.
- The Silhouette score metric obtains its best results with configurations that do not achieve good results in either the Davis-Bouldin index or the Calinski-Harabasz score. Moreover, this metric has a tendency to generate a large number of clusters.
- The number of components is the variable with the least impact on the clustering process. However, it is generally beneficial to apply a dimensionality reduction process to obtain the best results.

8.2 Media embeddings

Secondly, we are going to analyze the results obtained using embeddings generated from the captions. The model chosen to create these embeddings is ViT [39] for image and VideoMAE [42] for video. Both models were previously presented in Section 6. This model generates embeddings of size 768. All the results obtained using this type of embeddings, depending on the clustering algorithm used, can be seen in:

- DBSCAN: Results can be found in Table A.4.
- OPTICS: Results can be found in Table A.5.
- HDBSCAN: Results can be found in Table A.6.

8.2.1. DBSCAN

To streamline the presentation and examination of results, we have compiled the top-performing outcomes achieved through the utilization of the DBSCAN clustering algorithm in Table 8.4. We have categorized these configurations as the "best results," denoting those that yield the two highest results for each metric.

Neighbours	Components	Min samples	Davies-Bouldin Score	Calinski-Harabasz Score	Silhouette Score	Number of clusters
5	5	100	**2.9608**	93.7874	**0.0430**	3
5	3	100	3.0830	66.3080	-0.0555	5
100	3	50	4.1485	**156.5248**	-0.0194	6
100	3	100	4.0238	136.7877	0.0275	10
100	3	100	4.0238	136.7877	0.0275	10

Table 8.4: Top results using media embeddings and DBSCAN.

Regarding Table 8.4 we can observe how the Davis-Bould metric and the Silhouette score metric share hyper parameters in their best result. Both metrics use 5 as the number of neighbors 5 as the component name, and 5 as the minimum class size and generate 3 different classes. Regarding the Calinski-Harabasz metric, we can observe that its best result is achieved with 100 as the number of neighbors, 3 as the name of components, and 50 as the minimum class size. Observing the values of the metrics in general, we can see that the clusters created do not present a high quality dot com. This is because the values of the metric a are high compared to those obtained when using embeddings coming from text. The values of the Calinski-Harabasz metric are lower than those observed for two textual e, and the values of the Silhouette score are extremely close to zero, which indicates that the clusters are overlapping.

8.2.2. OPTICS

Table 8.5 shows the best results obtained by applying the OPTICS algorithm. We define the best results as those two parameter configurations that generate the best results for each metric.

Neighbours	Components	Min samples	Davies-Bouldin Score	Calinski-Harabasz Score	Silhouette Score	Number of clusters
100	3	100	**2.3814**	208.4453	**0.0844**	4
100	5	100	2.3836	208.3453	**0.0844**	4
100	50	500	3.1296	**232.4171**	0.0722	2
100	5	500	3.1199	232.2299	0.0723	2
100	50	100	2.3854	189.4246	0.0882	5
100	768	100	2.3955	208.2779	0.0845	4

Table 8.5: Top results using media embeddings and OPTICS.

Looking at Table 8.5 we can see that the best results for the Davis-Bouldin metric and the best results for the Silhouette score metric share a configuration; moreover, the best result for the Silhouette score metric can also be obtained with a configuration that obtains a performance almost equal to the best in the Davis-Bouldin score. These two configurations generate four different clusters and obtain a similar score in the Calinski-Harabasz metric. These two configurations share the values for two parameters, namely for the number of neighbors and for the minimum samples, using 100 as the value for the two parameters. Regarding the Calinski-Harabasz metric, we can observe that its best value is obtained using 500 as the minimum sample size, 50 as the number of components, and 100 as the number of neighbors, a very similar value and generates the same number of clusters (two) using 50 as the number of components. There is one constant throughout this table and that is the use of 100 as the ideal number of neighbors to achieve the best results. This is the variable with the greatest impact on the final quality of clusters when using embeddings from images and videos and the OPTICS algorithm.

8.2.3. HDBSCAN

In order to simplify the presentation of the results, we've extracted the most successful outcomes using HDBSCAN clustering algorithm to Table 8.3. The best results are defined as those configurations that generate the two best results for each metric.

Neighbours	Components	Min samples	Davies-Bouldin Score	Calinski-Harabasz Score	Silhouette Score	Number of clusters
50	768	100	**0.8445**	127.2098	**0.2216**	3
100	768	100	**0.8445**	127.2098	0.2214	2
50	3	50	0.9336	**253.8984**	0.2215	2
50	5	50	0.9336	**253.8984**	0.2215	2
50	5	100	0.9336	**253.8984**	0.2215	2

Table 8.6: Top results using media embeddings and HDBSCAN

In Table 8.6 we present the results obtained using the OPTICS algorithm. The first remarkable fact is that for the Silhouette score, the best configuration coincides with the best configuration for the Davis-Bouldin metric. In addition, there are two configurations that share the number of components equal to 768 and the minimum samples equal to 100 and obtain identical scores for the Davis-Bouldin metric and for the Calinski-Harabasz metric and extremely similar scores for the Calinski-Harabasz metric. Silhouette score; however, while the configuration using 50 as the number of neighbors generates 3 clusters, the configuration using 100 as the number of neighbors generates only two clusters. When we pay special attention to the Calinski-Harabasz metric, it is observable that the best result is obtained using three different configurations that generate the same results for the three metrics and the same number of clusters. These three configurations have in common to use 50 as the number of neighbors and an extremely low number of components, i.e., 3 or 5.

8.2.4. Media embeddings results

Paying attention to the conclusions drawn in the last three subsections, we can observe the following facts:

- When using embeddings coming from text, the Davis-Bouldin and Silhouette metrics obtain the best results using similar configurations. These configurations vary greatly depending on the clustering algorithm used but generate a similar number of clusters, generating 3 or 4 clusters.

- The DBSCAN and OPTICS algorithms are not suitable for unsupervised clustering with embeddings from images and videos as it does not achieve good scores on any of the three metrics.
- Dimensionality reduction is not a particularly important process when using media embeddings since its use does not greatly improve the results obtained by not performing dimensionality reduction.

8.3 Multimodal embeddings

Thirdly, we are going to analyze the results obtained using embeddings generated from media (image and video) and the captions. The model chosen to create these embeddings is ALIGN [43]. The model was previously presented in Section 6. This model generates embeddings of size 1280. All the results obtained using this type of embedding, depending on the clustering algorithm used, can be seen in the following:

- DBSCAN: Results can be found in Table A.7.
- OPTICS: Results can be found in Table A.8.
- HDBSCAN: Results can be found in Table A.9.

8.3.1. DBSCAN

To facilitate the presentation and assessment of the results, we've isolated the superior results attained through the utilization of the DBSCAN clustering algorithm, as detailed in Table 8.7. These optimal outcomes are characterized as the configurations yielding the two highest results for each metric.

Neighbours	Components	Min samples	Davies-Bouldin Score	Calinski-Harabasz Score	Silhouette Score	Number of clusters
100	3	500	**0.993**	**542.602**	0.062	2
100	5	50	2.047	236.379	0.244	20
100	1280	100	2.174	364.093	0.179	10
100	1280	50	2.135	241.184	**0.253**	21
100	50	50	2.104	252.178	0.251	20

Table 8.7: Top results using multimodal embeddings and DBSCAN.

The most characteristic fact present in Table 8.7 is that all the best performing configurations use 100 as the number of neighbors. When we pay attention to the metrics, we can observe that the Davis-Bouldin and Calinski-Harabasz metrics share the same configuration as the ideal one. This was already the case when text embeddings were used; however, it was not the case when embeddings were generated from images and videos. Paying attention to all the metrics, we observe a very remarkable feature: the configuration that generates the best results for the Davis-Bouldin and Calinski-Harabasz metrics is also the only one that generates a qualitatively analyzable number of clusters. This configuration shows very low results for the Silhouette score metric. Silhouette score; however, all the results that yield good values for the Silhouette score generate a large number of clusters that can be analyzed qualitatively. Silhouette score generates a large number of clusters and mediocre values for the other two metrics. Regarding the number of components used, we can observe how the Davis-Bouldin and Calinski-Harabasz metrics, in their best configuration, use only 3 components, while the Silhouette score does not support the use of dimensionality reduction techniques.

8.3.2. OPTICS

To streamline the presentation and examination of results, we have compiled the top-performing outcomes achieved through the utilization of the DBSCAN clustering algorithm in Table 8.8. We have categorized these configurations as the "best results," denoting those that yield the two highest results for each metric.

Neighbours	**Components**	**Min samples**	**Davies-Bouldin Score**	**Calinski-Harabasz Score**	**Silhouette Score**	**Number of clusters**
50	3	500	**0.990**	722.260	0.210	3
50	5	500	**0.990**	722.260	0.210	3
50	20	500	**0.990**	722.260	0.210	3
50	50	500	**0.990**	722.260	0.210	3
50	1280	500	**0.990**	722.260	0.210	3
5	5	1000	2.540	**799.930**	0.140	2
5	1280	1000	2.630	794.630	0.140	2
100	50	50	2.220	231.024	**0.250**	24
100	50	100	1.462	414.277	**0.250**	7

Table 8.8: Top results using multimodal embeddings and OPTICS.

Table 8.8 presents 5 different configurations to achieve the best results in the Davis-Bouldin metric. These have in common both the name of comma neighbors, which is 50, and the minimum samples, which is 500. The five configurations generate the same clusters, as can be observed by obtaining the same results in all metrics and generating the same number of clusters. This characteristic shows that the number of components is the variable that has the greatest impact on the clustering process. Regarding the Calinski-Harabasz metric, we can observe how its best configuration notably worsens the performance with respect to the Davis-Bouldin metric and obtains a mediocre value for the Silhouette score metric. The best Silhouette score programs generate a large number of clusters.

8.3.3. HDBSCAN

To enhance the organization and scrutiny of our results, we have extracted the most favorable outcomes produced by the HDBSCAN clustering algorithm and presented them in Table 8.9. We've defined these as the "best results," signifying configurations that produce the top two results for each metric.

Table 8.9 presents three different configurations to achieve the results of the point metric as well as using the OPTICS algorithm; the number of neighbors is 100, and the minimum samples is 500; also, the number of components is again a variable that does not affect the final clustering quality. Regarding the Calinski-Harabasz metric, it can be observed that the best configurations use a minimum clustering size of 500 and generate 3 clusters. With respect to the Silhouette score, the best configurations again generate a higher number of clusters and show a worsening with respect to the other metrics.

Neighbours	Components	Min samples	Davies-Bouldin Score	Calinski-Harabasz Score	Silhouette Score	Number of clusters
100	3	500	**0.991**	722.369	0.205	3
100	5	500	**0.991**	722.369	0.205	3
100	1280	500	**0.991**	722.369	0.205	3
5	3	500	2.341	**767.200**	0.177	3
50	5	500	1.645	726.461	0.202	3
50	3	100	2.157	428.877	**0.251**	7
50	20	100	1.159	418.769	0.249	7
50	1280	100	1.158	418.652	0.249	7
50	50	100	1.158	418.609	0.249	7

Table 8.9: Top results using multimodal embeddings and HDBSCAN.

However, it shows the same trend as that observed for the Davis-Bouldin metric, where the number of components is the variable that affects the clustering process the least.

8.3.4. Multimodal embeddings results

Paying attention to the conclusions drawn in the last three subsections, we can observe the following facts:

- Multimodal embeddings obtain competitive results in all three metrics used to evaluate the quality of the clustering process.
- The number of components used to perform the clustering process is the variable with the least impact on obtaining quality clusters.
- Good clusters can be obtained using very limited representation spaces.

In addition to these facts, the results obtained with multimodal embeddings present a pattern of special interest. Generally, after applying the clustering process, three clusters are obtained. This number is consistent with the intuitive idea that in a debate, there will be three main positions of people regarding the debate: there will be people in favor, people against, and people without a clear opinion.

8.4 General results

In order to be able to analyze in greater detail the results obtained previously and compare them in Table 8.10, we have recovered the results of the two best configurations for each metric together with some engagement configurations which, although they do not show the best results, obtain competitive results in the three metrics used. These compromise configurations are highlighted in italics.

Embedding	Algorithm	Neighbours	Components	Min cluster size	Davies-Bouldin Score	Calinski-Harabasz Score	Silhouette Score	Number of clusters
Media	HDBSCAN	50	768	100	**0.844**	127.209	0.221	3
Media	HDBSCAN	100	768	100	**0.844**	127.209	0.221	2
Multimodal	OPTICS	5	50	1000	**2.410**	**829.110**	0.150	2
Multimodal	OPTICS	5	20	1000	2.650	**807.180**	0.140	2
Multimodal	DBSCAN	100	1280	50	2.135	241.184	**0.253**	21
Multimodal	HDBSCAN	50	3	100	2.157	428.877	0.251	7
Multimodal	DBSCAN	100	50	50	2.104	252.178	0.251	20
Multimodal	*OPTICS*	*50*	*3*	*500*	*0.990*	*722.260*	*0.210*	*3*
Multimodal	*OPTICS*	*50*	*5*	*500*	*0.990*	*722.260*	*0.210*	*3*
Multimodal	*OPTICS*	*50*	*20*	*500*	*0.990*	*722.260*	*0.210*	*3*
Multimodal	*OPTICS*	*50*	*50*	*500*	*0.990*	*722.260*	*0.210*	*3*
Multimodal	*OPTICS*	*50*	*1280*	*500*	*0.990*	*722.260*	*0.210*	*3*
Multimodal	*HDBSCAN*	*100*	*3*	*500*	*0.991*	*722.369*	*0.205*	*3*
Multimodal	*HDBSCAN*	*100*	*5*	*500*	*0.991*	*722.369*	*0.205*	*3*
Multimodal	*HDBSCAN*	*100*	*1280*	*500*	*0.991*	*722.369*	*0.205*	*3*

Table 8.10: Best clustering results.

First of all, we will proceed to analyze the configurations that obtain the best performance by metric. There are two configurations that yield the best result for the Davis-Bouldin metric; these configurations are the only ones that appear in this table and that do not use multimodal embeddings; they only use embedding from videos and photographs. Although these two configurations obtain the best value in this metric, their performance in the Calinski-Harabasz metric is low. Moreover, these two configurations only differ in the number of neighbors. When we use 50 as the number of neighbors, we obtain 3 clusters; however, when using 100 neighbors, we recover only two clusters. Regarding the Calinski-Harabasz metric, we can observe that the two configurations that obtain the best performance use the OPTICS algorithm and multimodal embedding. These two configurations only differ in the number of components and generate two clusters; the weak point of these configurations is in the Davis-Bouldin and Silhouette score metrics. Silhouette score metrics since it obtains poor values in both of them. With respect to the Silhouette score, we can observe how there are two configurations that generate it, one using the HDBSCAN algorithm and the other using the DBSCAN algorithm; since each configuration is obtained by a different algorithm, the number of classes obtained is very different (and high), and the parameters used to generate them are also very different.

Secondly, we will analyze the compromise configurations, i.e., those that, although they do not obtain the best results in any metric, obtain competitive results in all of them. All the engagement configurations use multimodal embeddings, and none of them use the DBSCAN algorithm. All the compromise configurations use 500 as a minimum cluster, and the number of neighbors depends on which algorithm we are using. When we use the OPTICS algorithm, the number of neighbors to use is 50; when we use the HDBSCAN algorithm, the number of neighbors to use is 100. No matter which of the two algorithms we use, the number of components is the least significant variable since there are multiple configurations that obtain similar results varying this.

Once we have seen all these configurations, we conclude that it is advisable to use multimodal embeddings since they allow clustering to obtain high values for all the quality metrics and, at the same time, generate a number of clusters which is perfectly analyzable at a qualitative level. Regarding the algorithm, there are the OPTICS and HDBSCAN options, both of which generate good results depending on the parameters chosen. It is considered advisable to use the HDBSCAN algorithm since it has been slightly faster in the experimentation process. In order to study more deeply the generated clusters, we have decided to use multimodal embeddings together with the HDSCAN algorithm, projecting the data to a three-dimensional representation space. We have chosen this configuration because it allows us to perform 3D visualizations of the clusters while obtaining competitive results in the three metrics used.

CHAPTER 9
Cluster Analysis

In this section, we will proceed to study in depth the clusters generated from the experimentation carried out in Section 8. Since we are working with an unsupervised dataset, this task will be carried out from two perspectives: the polarity present in each cluster created and the named entities predominant in each cluster. The clustering process will be carried out by multimodal embeddings using the HDBSCAN algorithm. The optimal parameters have been studied in Section 8 and are as follows:

- **Number of neighbors:** 100. When performing the dimensionality reduction from the original representation space (1280 dimensions) to the final representation space (3 dimensions).
- **Number of components:** three. A three-dimensional representation space is used to perform the clustering process. One of the main advantages of utilizing such a low dimensionality representation space is that it will allow us to visualize the data.
- **Minimum cluster size:** 500. Each cluster will contain at least 500 data samples. The minimum number of data points that must be within a certain distance of a core point for those data points to be considered part of the same cluster.

9.0.1. Clustering characteristics

Each of the clusters created has different characteristics, as shown in Table 9.1.

	Number of posts	**Median of likes**	**Median of comments**	**Median of interactions**	**Median of hashtags**	**Median of mentions**	**Avg posts by profile**
Cluster 1	313	7	0	7	37	0	52.16
Cluster 2	4673	16	0	17	24	0	4.92
Cluster 3	879	4	0	4	31	0	97.66

Table 9.1: Cluster statistics

Since the distributions of the number of likes, comments, hashtags, and mentions are extremely unequal, we have decided to compare these statistics using the median.

Cluster 1 is the smallest cluster. Its posts are characterized by containing a large number of hashtags, and its users are quite active in posting about abortion on Instagram. The table 9.3 shows the ten most frequent hashtags in their publications and the number of times they appear. The ten most used hashtags appear more times than posts that exist

Hashtags	Number of occurrences
noalaborto	1243
encontradelaborto	930
asesinasaborteras	930
follow	930
like	930
fueraaborteras	930
nuncaafavor	620
fueralasproaborto	620
asesinas	620
fueraproasesinatos	620

Table 9.2: Ten most used hashtags in cluster 1.

within the cluster; this is because the posts contain the hashtag repeated multiple times. This fact also happens, although to a lesser extent, in the other clusters. This cluster uses an abundance of particularly aggressive hashtags and seeks to increase the reach of its message by using hashtags such as "#follow" or "#like".

Hashtags	Number of occurrences
noalaborto	4730
provida	2466
sialavida	2019
aborto	1238
conmishijosnotemetas	1036
feminismo	994
salvemoslasdosvidas	959
abortolegal	914
prolife	891
niunamenos	860

Table 9.3: Ten most used hashtags in cluster 2.

Cluster 2 contains most of the publications. Posts belonging to this cluster tend to receive a higher number of likes and are the ones that generally contain a lower number of hashtags. Regarding the most used hashtags, only one ("no abortion") appears more times than the total number of posts. Within this cluster we can observe both hashtags related to feminism and the defense of abortion rights and hashtags advocating the limitation of abortion rights. It is the cluster that uses more moderate hashtags, as cluster 1 uses some such as "#asesinas" and cluster 3 makes abundant use of hashtags that claim a position over time and space such as "#providaforever" or "#providamundial".

Cluster 3 groups the most active users, although it is the cluster that generally receives the fewest likes and interactions. In the publications of this cluster, there is a lower tendency to use hashtags repeatedly, as can be seen in Table 9.4; however, the repetition of hashtags is observed.

The clusters are not differentiable solely based on the most used hashtags, the number of publications per profile, or the reach they have. These three clusters are also separable in the representation space projected in the process, as can be seen in Figure 9.1.

Hashtags	Number of occurrences
providamundial	1756
soyprovida	1756
noalaborto	879
follow	878
brasil	878
provida	878
providaforever	878
sialavidanoalaborto	878
soloprovidas	878
somosprovidas	878

Table 9.4: Ten most used hashtags in cluster 3.

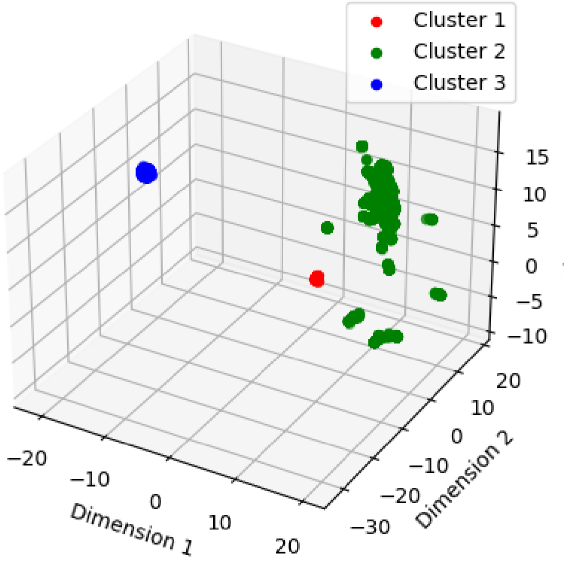


Figure 9.1: Named entity categories distribution.

9.0.2. Sentiment Analysis over clusters

In order to better understand the clusters created, we have analyzed the polarities present with the model previously used in Section 5.

Regarding cluster 1, we can observe that the majority of the posts have a neutral polarity, and only 3.51% of the publications have a negative polarity, as can be seen in Figure 9.2. None of the publications present a positive polarity. Although the large number of publications with neutral polarity may clash, it is important to remember that the neutral category contains both publications that do not present polarity and those that have mixed polarities, that is, those in which a positive and negative polarity are expressed. The large number of neutral posts is precisely due to this since cluster 1 is an aggressive cluster in which the publications do not hesitate to attack people who defend the right to abortion while defending its illegalization.

Cluster 2 is the cluster that contains the most moderate debate. Among the most used hashtags are hashtags that defend the right to abortion and hashtags that defend making access to this right difficult. This moderation and diversity of opinions can also be seen in Figure 9.3, which represents the polarity distribution of the publications. This is the unique cluster that contains all three polarities represented.

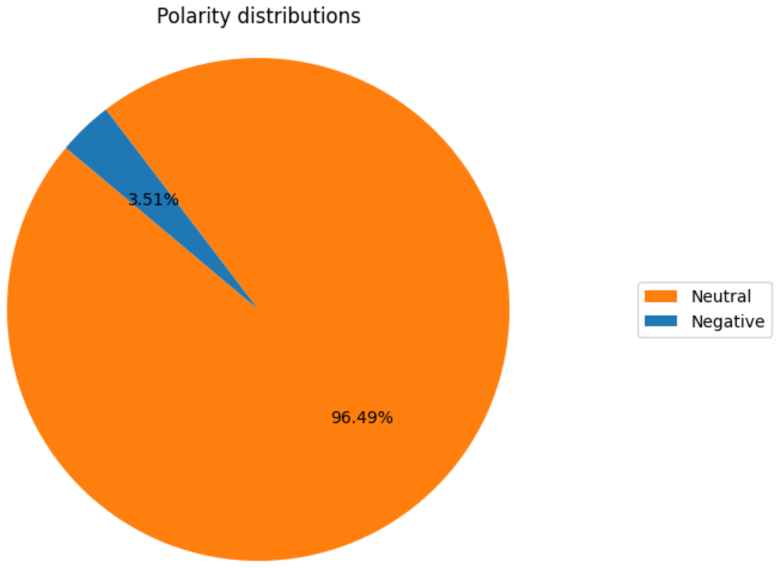


Figure 9.2: Distribution of the polarity present in the posts in cluster 1.

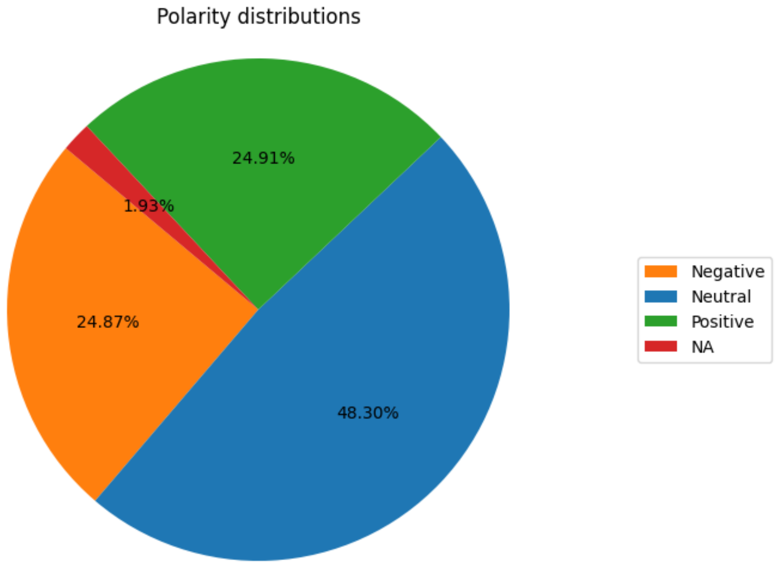


Figure 9.3: Distribution of the polarity present in the posts in cluster 2.

Cluster 3 is the cluster with the most activity by user. The publications belonging to this hashtag aim to increase the reach of the broadcasting accounts in order to be able to influence the debate about the right to abortion to a greater extent. As shown in Figure 9.4, the only polarity present in the messages of this cluster is negative.

9.0.3. Named entities over clusters

In order to better understand the clusters created, we have recognized the named entities present with the model previously used in Section 5.

Cluster 1 barely contains any mention of named entities about people, locations, or organizations, as shown in Figure 9.5. Although the publications of this cluster contain abundant attacks, these attacks are not made against entities but are launched in general.

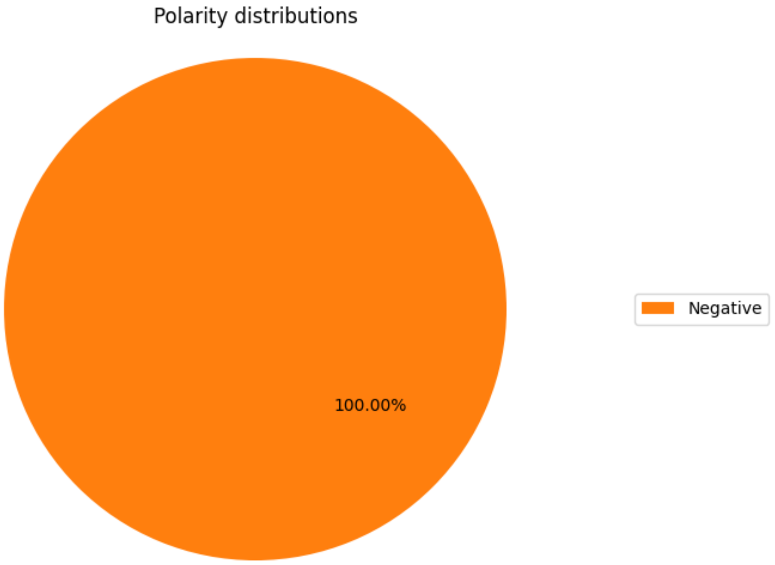


Figure 9.4: Distribution of the polarity present in the posts in cluster 3.

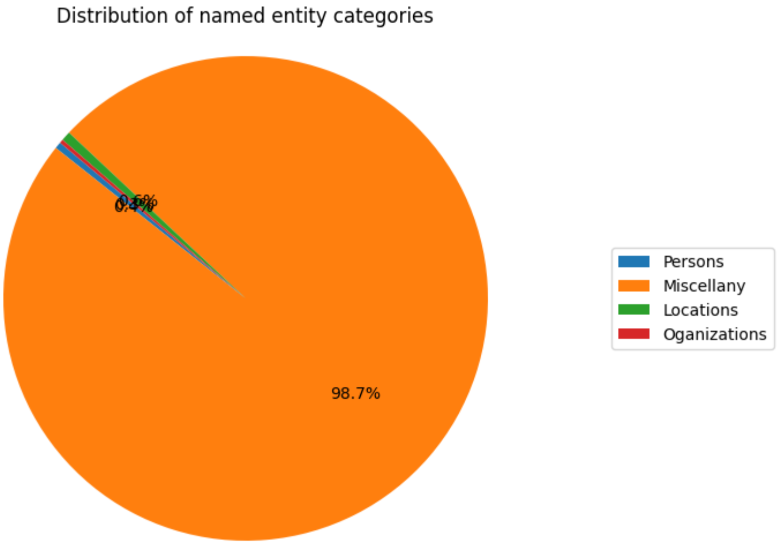


Figure 9.5: Named entity categories distribution.

The second cluster is characterized by grouping most of the debate and the most moderate part of it. This moderation and generalization of the debate can be seen in Figure 9.6. The debate becomes broader, and references are made to people, organizations, and locations. This last category is of special interest since the right to abortion is in different legal status depending on countries or regions.

Cluster 3 is the cluster where there are more publications per profile. This cluster brings together publications created with the direct intention of affecting public opinion on the abortion debate. Most of the named entities that appear in this cluster are related to locations. This is explained because the debate about the right to abortion is in different situations depending on the regions or countries where abortion is legalized and normalized to a greater extent, so they seek to mention the regions of interest to them with the aim of disseminating their ideas, especially in these regions.

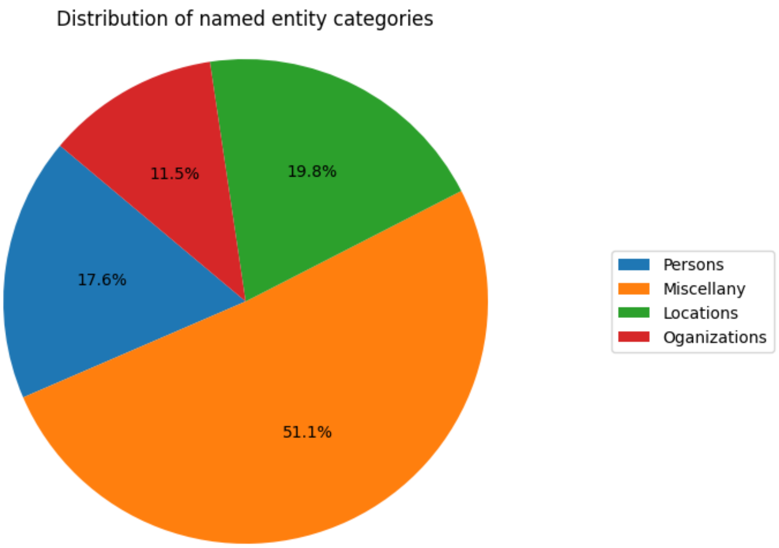


Figure 9.6: Named entity categories distribution.

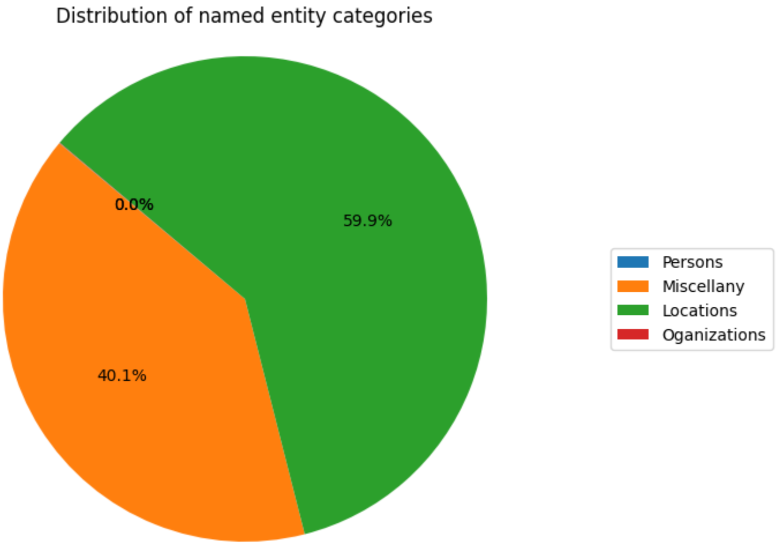


Figure 9.7: Named entity categories distribution.

CHAPTER 10

Conclusions and ongoing work

To conclude the work, the conclusions obtained from it will be presented, the achievement of objectives will be evaluated according to the rubric shown in Table 1.1, and the current lines of research and work on which we are currently working to extend this work.

10.1 Conclusions

This thesis has addressed all the necessary steps to approach the analysis of the abortion rights debate on Instagram.

To address this problem, we first created a new dataset containing Instagram posts (specifically multimedia content associated with the post, its caption, and different metadata) and proceeded to analyze it. Several analyses have been performed on this dataset, among which the analysis of languages used for the most active profiles and dates stands out. From these analyses that can be seen in Section 3, it is concluded that the debate on abortion rights is highly politicized, with many accounts explicitly created to disseminate publications to defend an ideological position. In addition, it is also observed that a large part of the debate is being treated from a Christian religious perspective with the influence of organized Christian organizations. In line with this, there is a significant increase in publications on days related to Christian festivities and commemorations.

The second step has been the creation of two frameworks that allow the analysis of debates on Instagram. These frameworks are presented in Section 4 . The first framework has been proposed that allows an exhaustive analysis of the dataset, facilitating the understanding of the available data and the debate, and a second reduced framework allows analysis with a reduced time cost. Both frameworks are agnostic to the artificial intelligence technologies used, thus allowing their adaptation to current and future state-of-the-art models.

The natural language processing tasks proposed in the extended framework were then carried out in Section 5. Firstly, the sentiment analysis task has been developed to analyze the publications' polarity. In this analysis, it has been observed that the number of publications with a positive polarity is low. Secondly, we proceeded to detect named entities. From the analysis of the named entities detected, many entities related to the Christian religion and ultra-right movements stand out. The large number of entities representing Latin American countries also stands out.

The next step was to analyze the techniques used to create embeddings in Section 6. It has been decided to experiment with three approaches: a first approach in which embeddings are generated only from the captions, a second approach in which only the captions are used, and a third approach in which the embeddings are generated only from the captions.

Once the multiple techniques used to create the embeddings were defined, we proceeded to study the preprocessing necessary to perform clustering and the techniques used in Section 6. Firstly, the UMAP technique has been presented to reduce the dimensionality of the embeddings to improve and accelerate the clustering process. The three clustering techniques that have been experimented with have been presented below. The techniques are DBSCAN, OPTICS, and HDBSCAN.

The next step in the work has been experimentation with different embedding and clustering techniques. From this experimentation, carried out in Section 8, it is concluded that multimodal representations generate higher-quality clusters than the other kinds of embeddings, media embeddings can be used to create embeddings although they lose performance, and textual embeddings are not suitable to create quality clusters.

Finally, the clusters generated by the configuration of greatest interest proposed in Section 9 have been analyzed. Three distinguishable clusters with well-defined patterns have been found. The first cluster found groups the most aggressive publications that try to increase the reach of the profiles that publish them, the second cluster contains the more moderate publications coming from the two sides of the debate, and the third cluster formed by those publications of the profiles that focus part of their participation in specific regions.

10.2 Evaluation of objectives

According to the rubric shown in Table 1.1, the objectives of the work have been achieved to varying degrees, as shown in the table 10.1

Objectives	**Objective not accomplished**	**Objective insufficiently accomplished**	**Objective sufficiently accomplished**	**Objective fully accomplished**
Creation of a dataset				✓
Definition and use of a user profiling framework at post level			✓	
Study of embedding tools to represent the information			✓	
Study of NLP tasks over the dataset			✓	
Paper publication		✓		

Table 10.1: Evaluation of objectives.

The dataset has been recovered and meets all the requirements set out at the beginning of the work, containing both the videos and photographs of the publications, the text present in the captions, and different metadata provided by Instagram.

Regarding the definition and use of multiple frameworks for the analysis of user profiles, we consider that the objective has been sufficiently met. However, we believe that these frameworks can be improved and defined in greater detail with expert help in the near future.

Multiple sources of embeddings have been studied in order to analyze which ones provide more information about the posts in order to classify and differentiate them, so this task is considered sufficiently accomplished.

We have carried out two tasks (sentiment analysis and named entities recognition) of Natural Language Processing both on the set of publications and as a function of the different clusters. However, we have only used one model for each task, so we consider the task sufficiently accomplished.

Regarding the steps required for the future publication of the evolution of this work, we have begun to take steps, but much remains to be done. We have studied several journals where publication could be carried out, and we consider of special interest the interdisciplinary journal Social Science Computer Review, which covers social science instructional and research applications of computing, as well as societal impacts of information technology. This journal stands in Q1 categories such as Computer Science, Interdisciplinary Applications (SCIE), Information Science & Library Science (SSCI), and Social Sciences, Interdisciplinary (SSCI).

10.3 Ongoing work

In this work, we have defined and studied the tools necessary to carry out an analysis of the debate about abortion on Instagram. However, these have only been the first steps in a more in-depth study. The steps we are currently taking to increase the depth and scope of the study are:

- **Dataset labeling:** Labeling the dataset: The first step we want to perform is to label the dataset with expert help.
- **Dataset expansion:** Expansion of the dataset: In order to be able to train our own models specialized in Instagram and to be able to apply fine-tuning in greater depth, it would be of interest to expand the dataset.
- **Creation of predefined classes:** With expert help, we want to create predefined categories of publications and study these categories.
- **Fine-tunning of models for Instagram:** To improve the quality of the models used, we would like to fine-tune with Instagram data.
- **Paper publication:** Publication of scientific articles with future advances.

Bibliography

[1] T. Markham, “Review essay: Social media, politics and protest,” *http://dx.doi.org/10.1177/0163443716665101*, vol. 38, pp. 946–957, 8 2016. [Online]. Available: https://journals.sagepub.com/doi/10.1177/0163443716665101

[2] C. Dadas, “Chapter 1. hashtag activism: The promise and risk of "attention",” *Social Writing/Social Media: Publics, Presentations, and Pedagogies*, pp. 17–36, 9 2020.

[3] H. Crandall and C. M. Cunningham, “Media ecology and hashtag activism: #kaleidoscope,” *Explorations in Media Ecology*, vol. 15, pp. 21–32, 3 2016. [Online]. Available: https://intellectdiscover.com/content/journals/10.1386/eme.15.1.21_1

[4] M. Daby, “Feminist mobilization and the abortion debate in latin america: Lessons from argentina,” *Politics & Gender*, vol. 18, pp. 359–393, 2022.

[5] A. Cioffi, C. Cecannecchia, L. Cipolloni, A. Santurro, and F. Cioffi, “The importance of the international community in protecting the right to abortion: The cases of malta and of the us supreme court,” *Healthcare 2023, Vol. 11, Page 520*, vol. 11, p. 520, 2 2023. [Online]. Available: https://www.mdpi.com/2227-9032/11/4/520/htmhttps://www.mdpi.com/2227-9032/11/4/520

[6] J. Nelson, “Feminism, human rights, and abortion debates in mexico,” *Journal of Women's History*, vol. 34, pp. 119–140, 6 2022. [Online]. Available: https://muse.jhu.edu/pub/1/article/856767

[7] D. Mulinari, “In green and white: Feminist struggles for abortion rights in argentina,” *Struggles for Reproductive Justice in the Era of Anti-Genderism and Religious Fundamentalism*, pp. 11–37, 2023. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-31260-1_2

[8] H. Armiwulan, “Rights to abortion, pro-choice vs. pro-life: Case of indonesia and the usa,” *International Journal of Criminal Justice Sciences*, vol. 17, pp. 128–139, 1 2022.

[9] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan Claypool Publishers, 2012. [Online]. Available: https://link.springer.com/10.1007/978-3-031-02145-9

[10] J. Ángel González Barba, “Attention-based approaches for text analytics in social media and automatic summarization,” 7 2021. [Online]. Available: https://riunet.upv.es/handle/10251/172245

[11] M. C. Díaz-Galiano, M. G. Vega, E. Casasola, L. Chiruzzo, M. Á. G. Cumbreras, E. M. Cámara, D. Moctezuma, A. Montejo-Ráez, M. A. S. Cabezudo, E. S. Tellez *et al.*, “Overview of tass 2019: One more further for the global spanish sentiment analysis corpus.” in *IberLEF@ SEPLN*, 2019, pp. 550–560.

[12] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review 2022 55:7*, vol. 55, pp. 5731–5780, 2 2022. [Online]. Available: https://link.springer.com/article/10.1007/s10462-022-10144-1

[13] B. Yang and C. Cardie, "Context-aware learning for sentence-level sentiment analysis with posterior regularization," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Baltimore, Maryland: Association for Computational Linguistics, Jun. 2014, pp. 325–335. [Online]. Available: https://aclanthology.org/P14-1031

[14] R. Sanchis-Font, M.-J. Castro-Bleda, B. Jorda-Albinana, and L. Lopez-Cuerva, "E-learning university evaluation through sentiment analysis centered on user experience dimensions," *DYNA*, vol. 98, no. 2, pp. 147–153, 2023.

[15] R. Sanchis-Font, M. J. Castro-Bleda, J. Ángel González, F. Pla, and L. F. Hurtado, "Cross-domain polarity models to evaluate user experience in e-learning," *Neural Processing Letters*, vol. 53, pp. 3199–3215, 10 2021. [Online]. Available: https://link.springer.com/article/10.1007/s11063-020-10260-5

[16] Z. B. Nezhad and M. A. Deihimi, "Twitter sentiment analysis from iran about covid 19 vaccine," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 16, no. 1, p. 102367, 2022.

[17] S. Verma, "Sentiment analysis of public services for smart society: Literature review and future research directions," *Government Information Quarterly*, vol. 39, no. 3, p. 101708, 2022.

[18] H. Li, Q. Chen, Z. Zhong, R. Gong, and G. Han, "E-word of mouth sentiment analysis for user behavior studies," *Information Processing & Management*, vol. 59, no. 1, p. 102784, 2022.

[19] K. Dashtipour, M. Gogate, A. Adeel, H. Larijani, and A. Hussain, "Sentiment analysis of persian movie reviews using deep learning," *Entropy 2021, Vol. 23, Page 596*, vol. 23, p. 596, 5 2021. [Online]. Available: https://www.mdpi.com/1099-4300/23/5/596/htmhttps://www.mdpi.com/1099-4300/23/5/596

[20] R. K. Behera, M. Jena, S. K. Rath, and S. Misra, "Co-lstm: Convolutional lstm model for sentiment analysis in social big data," *Information Processing Management*, vol. 58, p. 102435, 1 2021.

[21] P. Bhuvaneshwari, A. N. Rao, Y. H. Robinson, and M. N. Thippeswamy, "Sentiment analysis for user reviews using bi-lstm self-attention based cnn model," *Multimedia Tools and Applications*, vol. 81, pp. 12 405–12 419, 4 2022. [Online]. Available: https://link.springer.com/article/10.1007/s11042-022-12410-4

[22] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, pp. 4171–4186, 10 2018. [Online]. Available: https://arxiv.org/abs/1810.04805v2

[23] J. Ángel González, L. F. Hurtado, and F. Pla, "Twilbert: Pre-trained deep bidirectional transformers for spanish twitter," *Neurocomputing*, vol. 426, pp. 58–69, 2 2021.

[24] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, and P. G. Allen, "Roberta: A robustly optimized bert pretraining approach," 7 2019. [Online]. Available: https://arxiv.org/abs/1907.11692v1

[25] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 8440–8451, 11 2019. [Online]. Available: https://arxiv.org/abs/1911.02116v2

[26] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423

[27] A. CONNEAU and G. Lample, "Cross-lingual language model pretraining," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/c04c19c2c2474dbf5f7ac4372c5b9af1-Paper.pdf

[28] F. Barbieri, L. E. Anke, and J. Camacho-Collados, "Xlm-t: Multilingual language models in twitter for sentiment analysis and beyond," *2022 Language Resources and Evaluation Conference, LREC 2022*, pp. 258–266, 4 2021. [Online]. Available: https://arxiv.org/abs/2104.12250v2

[29] N. F. AL-Bakri, J. F. Yonan, and A. T. Sadiq, "Tourism companies assessment via social media using sentiment analysis," *Baghdad Science Journal*, vol. 19, no. 2, pp. 0422–0422, 2022.

[30] M. Tubishat, N. Idris, and M. Abushariah, "Explicit aspects extraction in sentiment analysis using optimal rules combination," *Future Generation Computer Systems*, vol. 114, pp. 448–480, 1 2021.

[31] B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," *Knowledge-Based Systems*, vol. 235, p. 107643, 1 2022.

[32] B. Mohit and I. Zitouni, "Named entity recognition," pp. 221–245, 2014. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-45358-8_7

[33] Y. Wang, H. Tong, Z. Zhu, and Y. Li, "Nested named entity recognition: A survey," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 16, 7 2022. [Online]. Available: https://dl.acm.org/doi/10.1145/3522593

[34] M. Ju, M. Miwa, and S. Ananiadou, "A neural layered model for nested named entity recognition," *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, pp. 1446–1459, 2018. [Online]. Available: https://aclanthology.org/N18-1131

[35] J. Yu, B. Bohnet, and M. Poesio, "Named entity recognition as dependency parsing," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 6470–6476, 2020. [Online]. Available: https://aclanthology.org/2020.acl-main.577

[36] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 7 2016. [Online]. Available: https://arxiv.org/abs/1607.04606v2

[37] B. Kantor and A. Globerson, “Coreference resolution with entity equalization,” *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pp. 673–677, 2019. [Online]. Available: https://aclanthology.org/P19-1066

[38] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, “A survey on vision transformer,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, pp. 87–110, 1 2023.

[39] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” *ICLR 2021 - 9th International Conference on Learning Representations*, 10 2020. [Online]. Available: https://arxiv.org/abs/2010.11929v2

[40] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin transformer: Hierarchical vision transformer using shifted windows,” *Proceedings of the IEEE International Conference on Computer Vision*, pp. 9992–10 002, 3 2021. [Online]. Available: https://arxiv.org/abs/2103.14030v2

[41] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-end object detection with transformers,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12346 LNCS, pp. 213–229, 5 2020. [Online]. Available: https://arxiv.org/abs/2005.12872v3

[42] Z. Tong, Y. Song, J. Wang, and L. Wang, “Videomae: Masked autoencoders are data-efficient learners for self-supervised video pre-training,” 3 2022. [Online]. Available: https://arxiv.org/abs/2203.12602v3

[43] C. Jia, Y. Yang, Y. Xia, Y. T. Chen, Z. Parekh, H. Pham, Q. V. Le, Y. Sung, Z. Li, and T. Duerig, “Scaling up visual and vision-language representation learning with noisy text supervision,” *Proceedings of Machine Learning Research*, vol. 139, pp. 4904–4916, 2 2021. [Online]. Available: https://arxiv.org/abs/2102.05918v2

[44] M. B. Sariyildiz, J. Perez, D. Larlus, M. B. Sariyildiz, J. Perez, and D. Larlus, “Learning visual representations with caption annotations,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12353 LNCS, pp. 153–170, 8 2020. [Online]. Available: https://arxiv.org/abs/2008.01392v1

[45] Y. Zhang, H. Jiang, Y. Miura, C. D. Manning, C. P. Langlotz, Y. Zhang, H. Jiang, Y. Miura, C. D. Manning, and C. P. Langlotz, “Contrastive learning of medical visual representations from paired images and text,” *Proceedings of Machine Learning Research*, vol. 182, pp. 1–24, 10 2020. [Online]. Available: https://arxiv.org/abs/2010.00747v2

[46] “Instaloader — download instagram photos and metadata.” [Online]. Available: https://instaloader.github.io/index.html

[47] “Shedding more light on how instagram works.” [Online]. Available: https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works

[48] "Mongodb: La plataforma de datos para aplicaciones | mongodb." [Online]. Available: https://www.mongodb.com/es

[49] "Cardiffnlp." [Online]. Available: https://cardiffnlp.github.io/

[50] M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd, "spaCy: Industrial-strength Natural Language Processing in Python," 2020. [Online]. Available: https://doi.org/10.5281/zenodo.1212303

[51] J. Nothman, N. Ringland, W. Radford, T. Murphy, and J. R. Curran, "Learning multilingual named entity recognition from wikipedia," *Artificial Intelligence*, vol. 194, pp. 151–175, 1 2013.

[52] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," *Journal of Machine Learning Research*, vol. 3, pp. 1137–1155, 8 2003.

[53] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," *8th International Conference on Learning Representations, ICLR 2020*, 9 2019. [Online]. Available: https://arxiv.org/abs/1909.11942v6

[54] M. Joshi, D. Chen, Y. Liu, D. S. Weld, L. Zettlemoyer, and O. Levy, "Spanbert: Improving pre-training by representing and predicting spans," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 64–77, 7 2019. [Online]. Available: https://arxiv.org/abs/1907.10529v3

[55] L. Wang, N. Yang, X. Huang, B. Jiao, L. Yang, D. Jiang, R. Majumder, and F. Wei, "Text embeddings by weakly-supervised contrastive pre-training," *arXiv preprint arXiv:2212.03533*, 2022.

[56] P. Bajaj, D. Campos, N. Craswell, L. Deng, J. Gao, X. Liu, R. Majumder, A. McNamara, B. Mitra, T. Nguyen, M. Rosenberg, X. Song, A. Stoica, S. Tiwary, and T. Wang, "Ms marco: A human generated machine reading comprehension dataset," *CEUR Workshop Proceedings*, vol. 1773, 11 2016. [Online]. Available: https://arxiv.org/abs/1611.09268v3

[57] V. Karpukhin, B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih, "Dense passage retrieval for open-domain question answering," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics, Nov. 2020, pp. 6769–6781. [Online]. Available: https://aclanthology.org/2020.emnlp-main.550

[58] T. Kwiatkowski, J. Palomaki, O. Redfield, M. Collins, A. Parikh, C. Alberti, D. Epstein, I. Polosukhin, J. Devlin, K. Lee, K. Toutanova, L. Jones, M. Kelcey, M.-W. Chang, A. M. Dai, J. Uszkoreit, Q. Le, and S. Petrov, "Natural questions: A benchmark for question answering research," *Transactions of the Association for Computational Linguistics*, vol. 7, pp. 452–466, 2019. [Online]. Available: https://aclanthology.org/Q19-1026

[59] N. Thakur, N. Reimers, A. Rücklé, A. Srivastava, and I. Gurevych, "Beir: A heterogenous benchmark for zero-shot evaluation of information retrieval models," 4 2021. [Online]. Available: https://arxiv.org/abs/2104.08663v4

[60] N. Muennighoff, N. Tazi, L. Magne, and N. Reimers, "Mteb: Massive text embedding benchmark," *EACL 2023 - 17th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, pp. 2006–2029, 10 2022. [Online]. Available: https://arxiv.org/abs/2210.07316v3

[61] C. Fields and C. Kennington, “Vision language transformers: A survey,” 7 2023. [Online]. Available: https://arxiv.org/abs/2307.03254v1

[62] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, “Object detection in 20 years: A survey,” *Proceedings of the IEEE*, vol. 111, pp. 257–276, 3 2023.

[63] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[64] T. Diwan, G. Anirudh, and J. V. Tembhurne, “Object detection using yolo: challenges, architectural successors, datasets and applications,” *Multimedia Tools and Applications*, vol. 82, pp. 9243–9275, 3 2023. [Online]. Available: https://link.springer.com/article/10.1007/s11042-022-13644-y

[65] R. Girshick, “Fast r-cnn,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.

[66] Z. Zhang and D. Tao, “Slow feature analysis for human action recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 436–450, 2012.

[67] W. Su, X. Zhu, Y. Cao, B. Li, L. Lu, F. Wei, and J. Dai, “Vl-bert: Pre-training of generic visual-linguistic representations,” *8th International Conference on Learning Representations, ICLR 2020*, 8 2019. [Online]. Available: https://arxiv.org/abs/1908.08530v4

[68] H. Tan and M. Bansal, “Lxmert: Learning cross-modality encoder representations from transformers,” *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, pp. 5100–5111, 8 2019. [Online]. Available: https://arxiv.org/abs/1908.07490v3

[69] A. Singh, R. Hu, V. Goswami, G. Couairon, W. Galuba, M. Rohrbach, and D. Kiela, “Flava: A foundational language and vision alignment model,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 15 638–15 650.

[70] J. Wang, D. Chen, Z. Wu, C. Luo, L. Zhou, Y. Zhao, Y. Xie, C. Liu, Y.-G. Jiang, and L. Yuan, “Omnivl: One foundation model for image-language and video-language tasks,” in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022, pp. 5696–5710. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/259a5df46308d60f8454bd4adcc3b462-Paper-Conference.pdf

[71] A. Diaz-Papkovich, L. Anderson-Trocmé, and S. Gravel, “A review of umap in population genetics,” *Journal of Human Genetics 2020 66:1*, vol. 66, pp. 85–91, 10 2020. [Online]. Available: https://www.nature.com/articles/s10038-020-00851-4

[72] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, “Dbscan revisited, revisited: Why and how you should (still) use dbscan,” *ACM Trans. Database Syst.*, vol. 42, no. 3, jul 2017. [Online]. Available: https://doi.org/10.1145/3068335

[73] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, “Optics: Ordering points to identify the clustering structure,” in *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD ’99. New York, NY,

USA: Association for Computing Machinery, 1999, p. 49–60. [Online]. Available: https://doi.org/10.1145/304182.304187

[74] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Advances in Knowledge Discovery and Data Mining*, J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 160–172.

[75] L. McInnes and J. Healy, "Accelerated hierarchical density based clustering," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017, pp. 33–42.

[76] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 11 1987.

[77] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.

[78] T. Caliński and H. JA, "A dendrite method for cluster analysis," *Communications in Statistics - Theory and Methods*, vol. 3, pp. 1–27, 01 1974.

APPENDIX A
Anexo A

A.1 Text embeddings results

Neighbors	Components	Min samples	Davies-Bouldin index	Calinski-Harabasz Index	Silhouette Coefficient	Number of clusters
5	3	50	3.3171	93.9676	0.1665	27
5	3	100	1.7356	368.1768	0.1247	4
5	3	500	N/A	N/A	N/A	1
5	3	1000	N/A	N/A	N/A	1
5	5	50	3.1760	92.7959	0.1592	27
5	5	100	1.7345	364.0150	0.1220	4
5	5	500	N/A	N/A	N/A	1
5	5	1000	N/A	N/A	N/A	1
5	20	50	3.4872	86.9618	0.1718	30
5	20	100	1.8158	336.9284	0.1088	4
5	20	500	N/A	N/A	N/A	1
5	20	1000	N/A	N/A	N/A	1
5	50	50	3.2292	87.7655	0.1727	31
5	50	100	1.7777	354.4232	0.1132	4
5	50	500	N/A	N/A	N/A	1
5	50	1000	N/A	N/A	N/A	1
5	384	50	3.0487	96.2267	0.1474	25
5	384	100	1.7835	361.3581	0.1232	4
5	384	500	N/A	N/A	N/A	1
5	384	1000	N/A	N/A	N/A	1
50	3	50	2.3719	158.0386	0.2078	17
50	3	100	3.0157	222.5289	0.1712	10
50	3	500	**1.1071**	**736.7043**	0.1140	2
50	3	1000	N/A	N/A	N/A	1
50	5	50	2.4558	166.1708	0.2060	16
50	5	100	3.1859	205.6347	0.1746	11
50	5	500	1.1073	732.4475	0.1128	2
50	5	1000	N/A	N/A	N/A	1
50	20	50	2.4832	151.4843	0.2073	18
50	20	100	3.0516	221.7901	0.1725	10
50	20	500	1.1080	718.6709	0.1088	2

50	20	1000	N/A	N/A	N/A	1
50	50	50	2.4697	158.8645	0.2072	17
50	50	100	3.1371	205.5075	0.1742	11
50	50	500	**1.1071**	719.8654	0.1092	2
50	50	1000	N/A	N/A	N/A	1
50	384	50	2.4691	158.4751	0.2062	17
50	384	100	3.1518	204.7188	0.1735	11
50	384	500	1.1076	717.7036	0.1085	2
50	384	1000	N/A	N/A	N/A	1
100	3	50	2.3730	161.0999	0.2024	17
100	3	100	2.9929	261.4415	0.1636	8
100	3	500	1.1093	725.8035	0.1107	2
100	3	1000	N/A	N/A	N/A	1
100	5	50	2.4574	156.4663	**0.2097**	18
100	5	100	3.2105	211.4509	0.1757	11
100	5	500	1.1079	732.3344	0.1127	2
100	5	1000	N/A	N/A	N/A	1
100	20	50	2.4262	152.8972	0.2046	18
100	20	100	3.1617	209.8641	0.1747	11
100	20	500	1.1114	711.8630	0.1065	2
100	20	1000	N/A	N/A	N/A	1
100	50	50	2.4722	155.9141	0.2095	18
100	50	100	3.1729	207.4289	0.1735	11
100	50	500	1.1119	712.8273	0.1067	2
100	50	1000	N/A	N/A	N/A	1
100	384	50	2.4599	154.1572	0.2062	18
100	384	100	3.2062	210.6160	0.1754	11
100	384	500	1.1162	685.3186	0.0982	2
100	384	1000	N/A	N/A	N/A	1

Table A.1: DBSCAN results using text embeddings.

Neighbors	Components	Min samples	Davies-Bouldin index	Calinski-Harabasz Index	Silhouette Coefficient	Number of clusters
5	3	50	3.1853	88.5346	0.1807	30
5	3	100	2.6832	195.8012	0.1519	10
5	3	500	4.2694	450.1741	0.1215	3
5	3	1000	4.1748	323.5483	0.0807	2
5	5	50	3.1058	93.5020	0.1774	28
5	5	100	2.9858	174.2845	0.1405	10
5	5	500	3.3015	467.2811	0.1286	3
5	5	1000	N/A	N/A	N/A	1
5	20	50	2.9628	94.9918	0.1790	29
5	20	100	3.3012	201.3895	0.1576	10
5	20	500	3.0087	476.9794	0.1298	3
5	20	1000	4.0815	334.8827	0.0815	2
5	50	50	3.3386	98.0605	0.1745	25
5	50	100	2.5941	204.9264	0.1417	9
5	50	500	3.8177	466.5900	0.1250	3
5	50	1000	3.2613	456.4139	0.0908	2
5	384	50	3.1158	94.4657	0.1790	29
5	384	100	3.3635	231.4516	0.1385	8
5	384	500	3.8568	464.2851	0.1249	3
5	384	1000	4.6396	262.2251	0.0763	2
50	3	50	2.2710	150.9402	0.2093	19
50	3	100	2.0270	230.8284	0.1668	9
50	3	500	4.0485	463.9867	0.1290	3
50	3	1000	2.1783	610.8564	0.0919	2
50	5	50	2.2260	141.0449	0.2096	20
50	5	100	2.2438	205.2416	0.1646	10
50	5	500	4.0312	460.5883	0.1285	3
50	5	1000	N/A	N/A	N/A	1
50	20	50	2.0698	152.3755	0.2058	18
50	20	100	1.9575	235.3262	0.1666	9
50	20	500	4.0338	460.3673	0.1285	3
50	20	1000	N/A	N/A	N/A	1
50	50	50	2.2060	149.4639	0.2189	20
50	50	100	2.8121	257.5974	0.1856	9
50	50	500	4.0416	464.0547	0.1290	3
50	50	1000	N/A	N/A	N/A	1
50	384	50	2.2455	148.3657	0.2158	20
50	384	100	1.8600	258.2111	0.1652	8
50	384	500	4.0402	464.1154	0.1291	3
50	384	1000	2.0104	625.2015	0.0979	2
100	3	50	2.0040	151.8250	0.2044	18
100	3	100	2.1309	230.6781	0.1647	9
100	3	500	1.1845	730.7169	0.1139	2
100	3	1000	N/A	N/A	N/A	1
100	5	50	1.8589	88.7060	-0.0614	17
100	5	100	1.8695	292.8080	0.1835	8
100	5	500	2.3586	506.9571	0.1443	3

100	5	1000	N/A	N/A	N/A	1
100	20	50	1.9777	93.2880	-0.0654	18
100	20	100	2.0924	230.7519	0.1671	9
100	20	500	1.1845	730.7169	0.1139	2
100	20	1000	N/A	N/A	N/A	1
100	50	50	2.0182	139.9196	0.2016	19
100	50	100	1.8775	276.7891	0.1745	8
100	50	500	1.1845	730.7169	0.1139	2
100	50	1000	N/A	N/A	N/A	1
100	384	50	1.9874	138.1916	0.1901	18
100	384	100	1.8641	255.3424	0.1615	8
100	384	500	3.9936	461.2582	0.1287	3
100	384	1000	N/A	N/A	N/A	1

Table A.2: OPTICS results using text embeddings.

Neighbors	Components	Min samples	Davies-Bouldin index	Calinski-Harabasz Index	Silhouette Coefficient	Number of clusters
5	3	50	3.185	88.535	0.181	30
5	3	100	2.683	195.801	0.152	10
5	3	500	3.371	465.029	0.128	3
5	3	1000	N/A	N/A	N/A	1
5	5	50	2.909	148.889	0.160	16
5	5	100	2.788	208.579	0.145	9
5	5	500	2.768	474.504	0.129	3
5	5	1000	N/A	N/A	N/A	1
5	20	50	2.956	142.772	0.173	17
5	20	100	2.488	252.484	0.140	7
5	20	500	2.743	473.792	0.131	3
5	20	1000	N/A	N/A	N/A	1
5	50	50	2.909	140.569	0.185	18
5	50	100	3.248	235.563	0.142	8
5	50	500	2.862	467.277	0.133	3
5	50	1000	N/A	N/A	N/A	1
5	384	50	3.091	117.323	0.159	20
5	384	100	2.832	268.289	0.141	7
5	384	500	3.189	448.944	0.127	3
5	384	1000	N/A	N/A	N/A	1
50	3	50	2.486	159.618	0.213	18
50	3	100	3.221	263.920	0.190	9
50	3	500	1.437	426.386	0.080	3
50	3	1000	N/A	N/A	N/A	1
50	5	50	2.419	167.993	0.214	17
50	5	100	3.059	253.648	0.183	9
50	5	500	1.413	420.661	0.078	3
50	5	1000	N/A	N/A	N/A	1
50	20	50	2.176	170.532	0.208	16
50	20	100	2.965	228.630	0.183	10
50	20	500	1.189	729.444	0.114	2
50	20	1000	N/A	N/A	N/A	1
50	50	50	2.421	168.964	0.218	17
50	50	100	2.910	237.011	0.189	10
50	50	500	2.152	490.869	0.128	3
50	50	1000	N/A	N/A	N/A	1
50	384	50	2.450	167.801	0.215	17
50	384	100	2.933	224.427	0.178	10
50	384	500	3.189	448.944	0.127	3
50	384	1000	N/A	N/A	N/A	1
100	3	50	2.363	167.172	0.215	17
100	3	100	2.942	252.309	0.181	9
100	3	500	1.188	729.958	0.114	2
100	3	1000	N/A	N/A	N/A	1
100	5	50	2.400	165.796	0.211	17
100	5	100	3.136	263.293	0.188	9
100	5	500	1.186	730.088	0.114	2

100	5	1000	N/A	N/A	N/A	1
100	20	50	2.403	157.983	0.214	18
100	20	100	2.955	255.197	0.181	9
100	20	500	1.185	730.717	0.114	2
100	20	1000	N/A	N/A	N/A	1
100	50	50	2.400	166.866	0.214	17
100	50	100	3.036	252.504	0.180	9
100	50	500	1.187	730.325	0.114	2
100	50	1000	N/A	N/A	N/A	1
100	384	50	2.670	156.194	0.210	18
100	384	100	2.933	224.427	0.178	10
100	384	500	3.189	448.944	0.127	3
100	384	1000	N/A	N/A	N/A	1

Table A.3: HDBSCAN results using text embeddings.

A.2 Media embeddings results

Neighbors	Components	Min samples	Davies-Bouldin index	Calinski-Harabasz Index	Silhouette Coefficient	Number of clusters
5	3	50	3.7521	66.0827	-0.0211	22
5	3	100	3.0830	66.3080	-0.0555	5
5	3	500	N/A	N/A	N/A	1
5	3	1000	N/A	N/A	N/A	1
5	5	50	3.5704	59.4787	-0.0360	23
5	5	100	2.9608	93.7874	0.0430	3
5	5	500	N/A	N/A	N/A	1
5	5	1000	N/A	N/A	N/A	1
5	20	50	3.5262	59.6084	-0.0422	23
5	20	100	3.9691	61.3205	-0.0625	5
5	20	500	N/A	N/A	N/A	1
5	20	1000	N/A	N/A	N/A	1
5	50	50	3.6769	64.4495	-0.0368	21
5	50	100	3.5615	71.1538	-0.0386	4
5	50	500	N/A	N/A	N/A	1
5	50	1000	N/A	N/A	N/A	1
5	768	50	3.7441	46.7454	-0.0544	22
5	768	100	3.8203	61.5561	-0.0687	5
5	768	500	N/A	N/A	N/A	1
5	768	1000	N/A	N/A	N/A	1
50	3	50	4.1631	134.9706	-0.0144	8
50	3	100	3.6251	120.5523	0.0179	12
50	3	500	N/A	N/A	N/A	1
50	3	1000	N/A	N/A	N/A	1
50	5	50	4.3526	114.1221	-0.0147	10
50	5	100	3.3129	121.4638	0.0120	12
50	5	500	N/A	N/A	N/A	1
50	5	1000	N/A	N/A	N/A	1
50	20	50	4.4460	113.2306	-0.0157	10
50	20	100	3.2933	120.1458	0.0135	12
50	20	500	N/A	N/A	N/A	1
50	20	1000	N/A	N/A	N/A	1
50	50	50	4.0667	107.8771	-0.0122	12
50	50	100	3.2977	113.9052	0.0014	13
50	50	500	N/A	N/A	N/A	1
50	50	1000	N/A	N/A	N/A	1
50	768	50	4.3626	102.7456	-0.0314	11
50	768	100	3.3001	126.0479	0.0194	11
50	768	500	N/A	N/A	N/A	1
50	768	1000	N/A	N/A	N/A	1
100	3	50	4.1485	156.5248	-0.0194	6
100	3	100	4.0238	136.7877	0.0275	10
100	3	500	N/A	N/A	N/A	1
100	3	1000	N/A	N/A	N/A	1

100	5	50	4.5051	123.2353	-0.0026	9
100	5	100	3.5264	124.0507	0.0191	12
100	5	500	N/A	N/A	N/A	1
100	5	1000	N/A	N/A	N/A	1
100	20	50	4.4337	115.0651	-0.0099	10
100	20	100	3.3598	132.6871	0.0236	11
100	20	500	N/A	N/A	N/A	1
100	20	1000	N/A	N/A	N/A	1
100	50	50	4.5107	115.7895	-0.0039	10
100	50	100	3.3558	127.9906	0.0241	12
100	50	500	N/A	N/A	N/A	1
100	50	1000	N/A	N/A	N/A	1
100	768	50	4.3874	105.5987	-0.0172	11
100	768	100	3.3895	132.4437	0.0225	11
100	768	500	N/A	N/A	N/A	1
100	768	1000	N/A	N/A	N/A	1

Table A.4: DBSCAN results using media embeddings.

Neighbors	Components	Min samples	Davies-Bouldin index	Calinski-Harabasz Index	Silhouette Coefficient	Number of clusters
5	3	50	3.5743	41.8272	-0.0674	22
5	3	100	3.1295	144.2070	-0.0017	6
5	3	500	3.3808	216.5157	0.0651	2
5	3	1000	N/A	N/A	N/A	1
5	5	50	3.4171	42.0360	-0.0781	18
5	5	100	3.1672	146.5050	0.0151	7
5	5	500	3.3011	221.8050	0.0665	2
5	5	1000	N/A	N/A	N/A	1
5	20	50	3.2807	53.5254	-0.0579	13
5	20	100	3.3444	109.1086	-0.0074	8
5	20	500	3.4033	229.4984	0.0701	2
5	20	1000	N/A	N/A	N/A	1
5	50	50	3.3702	50.6960	-0.0413	15
5	50	100	3.1804	116.8614	0.0109	9
5	50	500	3.1925	224.4356	0.0699	2
5	50	1000	N/A	N/A	N/A	1
5	768	50	3.3377	36.9040	-0.0795	14
5	768	100	3.3780	88.2315	-0.0352	6
5	768	500	3.2644	221.8985	0.0684	2
5	768	1000	N/A	N/A	N/A	1
50	3	50	2.7465	100.2212	-0.0240	10
50	3	100	2.7315	150.5992	0.0003	6
50	3	500	N/A	N/A	N/A	1
50	3	1000	N/A	N/A	N/A	1
50	5	50	2.6636	104.0081	-0.0302	10
50	5	100	2.7300	137.5864	-0.0110	7
50	5	500	3.1271	231.4301	0.0719	2
50	5	1000	N/A	N/A	N/A	1
50	20	50	2.7137	105.8851	-0.0188	10
50	20	100	2.6499	156.0027	0.0115	6
50	20	500	3.1326	231.5535	0.0723	2
50	20	1000	N/A	N/A	N/A	1
50	50	50	2.6825	75.8564	-0.0532	10
50	50	100	2.6007	158.3461	0.0187	6
50	50	500	3.1263	231.6378	0.0721	2
50	50	1000	N/A	N/A	N/A	1
50	768	50	2.5949	112.5937	-0.0205	9
50	768	100	2.9141	139.7358	0.0191	7
50	768	500	3.1263	231.6378	0.0721	2
50	768	1000	N/A	N/A	N/A	1
100	3	50	2.5935	87.5269	-0.0730	10
100	3	100	2.3814	208.4453	0.0844	4
100	3	500	N/A	N/A	N/A	1
100	3	1000	N/A	N/A	N/A	1
100	5	50	2.5386	97.4449	-0.0543	11
100	5	100	2.3836	208.3453	0.0844	4
100	5	500	3.1199	232.2299	0.0723	2

100	5	1000	N/A	N/A	N/A	1
100	20	50	2.7069	104.0152	-0.0188	10
100	20	100	2.6499	156.0027	0.0115	6
100	20	500	3.1326	231.5535	0.0723	2
100	20	1000	N/A	N/A	N/A	1
100	50	50	2.5714	124.6652	-0.0102	8
100	50	100	2.3854	189.4246	0.0882	5
100	50	500	3.1296	232.4171	0.0722	2
100	50	1000	N/A	N/A	N/A	1
100	768	50	2.7593	101.9772	-0.0101	11
100	768	100	2.3955	208.2779	0.0845	4
100	768	500	3.1263	231.6378	0.0721	2
100	768	1000	N/A	N/A	N/A	1

Table A.5: OPTICS results using media embeddings.

Neighbors	Components	Min samples	Davies-Bouldin index	Calinski-Harabasz Index	Silhouette Coefficient	Number of clusters
5	3	50	3.1853	88.5346	0.1807	30
5	3	100	2.6832	195.8012	0.1519	10
5	3	500	5.6746	154.0028	0.0495	3
5	3	1000	N/A	N/A	N/A	1
5	5	50	4.6223	131.7122	0.0313	5
5	5	100	4.3100	130.8371	0.0431	3
5	5	500	5.5707	165.4117	0.0519	3
5	5	1000	N/A	N/A	N/A	1
5	20	50	4.0370	168.0913	0.0347	4
5	20	100	5.5598	161.7902	0.0395	4
5	20	500	5.7830	166.5814	0.0417	3
5	20	1000	N/A	N/A	N/A	1
5	50	50	4.3174	176.6520	0.0339	4
5	50	100	3.5364	106.2773	0.0349	3
5	50	500	5.1374	178.7786	0.0635	3
5	50	1000	2.6855	2.5408	-0.1067	3
5	768	50	3.7348	40.1691	-0.0606	24
5	768	100	4.8662	113.2667	-0.0298	4
5	768	500	5.4695	171.8327	0.0507	3
5	768	1000	N/A	N/A	N/A	1
50	3	50	0.9336	253.8984	0.2215	2
50	3	100	1.2538	218.9760	0.1959	2
50	3	500	5.5092	164.7844	0.0449	3
50	3	1000	N/A	N/A	N/A	1
50	5	50	0.9336	253.8984	0.2215	2
50	5	100	0.9336	253.8984	0.2215	2
50	5	500	5.5502	168.0939	0.0459	3
50	5	1000	N/A	N/A	N/A	1
50	20	50	0.9336	253.8984	0.2215	2
50	20	100	0.9336	253.8984	0.2215	2
50	20	500	5.5569	167.6218	0.0444	3
50	20	1000	N/A	N/A	N/A	1
50	50	50	0.9336	253.8984	0.2215	2
50	50	100	0.9336	253.8984	0.2215	2
50	50	500	5.6349	165.3637	0.0396	3
50	50	1000	3.7129	1.8768	-0.0669	3
50	768	50	0.9336	253.8984	0.2215	2
50	768	100	0.8445	127.2098	0.2216	3
50	768	500	5.5131	169.0023	0.046	3
50	768	1000	N/A	N/A	N/A	1
100	3	50	0.9336	253.8984	0.2215	2
100	3	100	0.9336	253.8984	0.2215	2
100	3	500	5.1861	176.5335	0.0523	3
100	3	1000	N/A	N/A	N/A	1
100	5	50	0.9336	253.8984	0.2215	2
100	5	100	1.2538	218.9760	0.1959	2
100	5	500	5.3711	172.2859	0.0485	3

100	5	1000	N/A	N/A	N/A	1
100	20	50	0.9336	253.8984	0.2215	2
100	20	100	0.9336	253.8984	0.2215	2
100	20	500	5.2980	173.2517	0.0511	3
100	20	1000	N/A	N/A	N/A	1
100	50	50	0.9336	253.8984	0.2215	2
100	50	100	0.9336	253.8984	0.2215	2
100	50	500	5.4165	171.4731	0.0479	3
100	50	1000	0.9336	253.8984	0.2215	2
100	768	50	0.9336	253.8984	0.2215	2
100	768	100	0.8445	127.2098	0.2214	2
100	768	500	5.3602	172.3230	0.04960	3
100	768	1000	N/A	N/A	N/A	1

Table A.6: HDBSCAN results using media embeddings.

A.3 Multimodal embeddings

Neighbors	Components	Min samples	Davies-Bouldin index	Calinski-Harabasz Index	Silhouette Coefficient	Number of clusters
5	3	50	3.185	88.535	0.181	30
5	3	100	N/A	N/A	N/A	1
5	3	500	N/A	N/A	N/A	1
5	3	1000	N/A	N/A	N/A	1
5	5	50	2.383	91.443	-0.027	23
5	5	100	N/A	N/A	N/A	1
5	5	500	N/A	N/A	N/A	1
5	5	1000	N/A	N/A	N/A	1
5	20	50	2.615	91.779	-0.008	18
5	20	100	N/A	N/A	N/A	1
5	20	500	N/A	N/A	N/A	1
5	20	1000	N/A	N/A	N/A	1
5	50	50	2.350	90.847	-0.058	22
5	50	100	N/A	N/A	N/A	1
5	50	500	N/A	N/A	N/A	1
5	50	1000	N/A	N/A	N/A	1
5	1280	50	2.430	98.957	0.008	22
5	1280	100	N/A	N/A	N/A	1
5	1280	500	N/A	N/A	N/A	1
5	1280	1000	N/A	N/A	N/A	1
50	3	50	2.096	238.186	0.234	19
50	3	100	2.178	360.811	0.177	10
50	3	500	N/A	N/A	N/A	1
50	3	1000	N/A	N/A	N/A	1
50	5	50	2.171	243.877	0.244	21
50	5	100	2.161	341.614	0.156	10
50	5	500	N/A	N/A	N/A	1
50	5	1000	N/A	N/A	N/A	1
50	20	50	2.159	236.056	0.240	21
50	20	100	2.159	343.553	0.158	10
50	20	500	N/A	N/A	N/A	1
50	20	1000	N/A	N/A	N/A	1
50	50	50	2.163	229.754	0.230	21
50	50	100	2.140	294.532	0.090	10
50	50	500	N/A	N/A	N/A	1
50	50	1000	N/A	N/A	N/A	1
50	1280	50	2.174	225.813	0.248	22
50	1280	100	2.174	352.985	0.152	9
50	1280	500	N/A	N/A	N/A	1
50	1280	1000	N/A	N/A	N/A	1
100	3	50	2.070	234.825	0.242	20
100	3	100	2.211	327.506	0.179	12
100	3	500	0.993	542.602	0.062	2
100	3	1000	N/A	N/A	N/A	1

100	5	50	2.047	236.379	0.244	20
100	5	100	2.158	346.845	0.180	11
100	5	500	N/A	N/A	N/A	1
100	5	1000	N/A	N/A	N/A	1
100	20	50	2.098	232.393	0.242	20
100	20	100	2.222	326.972	0.181	12
100	20	500	N/A	N/A	N/A	1
100	20	1000	N/A	N/A	N/A	1
100	50	50	2.104	252.178	0.251	20
100	50	100	2.201	336.271	0.175	11
100	50	500	N/A	N/A	N/A	1
100	50	1000	N/A	N/A	N/A	1
100	1280	50	2.135	241.184	0.253	21
100	1280	100	2.174	364.093	0.179	10
100	1280	500	N/A	N/A	N/A	1
100	1280	1000	N/A	N/A	N/A	1

Table A.7: DBSCAN results using multimodal embeddings.

Neighbors	Components	Min samples	Davies-Bouldin index	Calinski-Harabasz Index	Silhouette Coefficient	Number of clusters
5	3	50	3.81	88.53	0.18	30
5	3	100	2.68	195.80	0.15	10
5	3	500	2.00	567.28	0.09	3
5	3	1000	N/A	N/A	N/A	1
5	5	50	2.21	105.30	-0.20	25
5	5	100	2.08	266.08	0.12	9
5	5	500	2.44	685.63	0.13	3
5	5	1000	2.54	799.93	0.14	2
5	20	50	2.31	129.58	-0.12	26
5	20	100	1.95	193.12	-0.01	8
5	20	500	1.96	663.08	0.13	3
5	20	1000	2.65	807.18	0.14	2
5	50	50	2.58	130.17	0.07	22
5	50	100	2.13	284.42	0.13	10
5	50	500	1.99	634.70	0.12	3
5	50	1000	2.41	829.11	0.15	2
5	1280	50	2.11	95.99	-0.04	20
5	1280	100	2.18	228.06	0.03	7
5	1280	500	2.57	589.99	0.11	3
5	1280	1000	2.63	794.63	0.14	2
50	3	50	1.94	161.44	0.12	22
50	3	100	1.69	339.74	0.21	9
50	3	500	0.99	722.26	0.21	3
50	3	1000	2.03	947.73	0.17	2
50	5	50	1.96	103.26	-0.08	22
50	5	100	1.82	392.74	0.21	10
50	5	500	0.99	722.26	0.21	3
50	5	1000	N/A	N/A	N/A	1
50	20	50	1.94	104.34	-0.09	22
50	20	100	1.51	386.18	0.17	8
50	20	500	0.99	722.26	0.21	3
50	20	1000	N/A	N/A	N/A	1
50	50	50	1.87	205.72	0.21	21
50	50	100	1.73	327.95	0.17	10
50	50	500	0.99	722.26	0.21	3
50	50	1000	N/A	N/A	N/A	1
50	1280	50	1.85	203.65	0.20	21
50	1280	100	1.15	418.86	0.25	7
50	1280	500	0.99	722.26	0.21	3
50	1280	1000	N/A	N/A	N/A	1
100	3	50	2.2085	200.6583	0.2289	26
100	3	100	1.6535	411.4413	0.2475	7
100	3	500	0.9920	722.3548	0.2050	3
100	3	1000	N/A	N/A	N/A	1
100	5	50	1.6452	249.8748	0.2090	12
100	5	100	1.5521	423.3533	0.2280	7
100	5	500	0.9913	722.3694	0.2051	3

100	5	1000	N/A	N/A	N/A	1
100	20	50	1.5358	249.8548	0.2093	12
100	20	100	1.2683	424.0352	0.2331	7
100	20	500	0.9920	722.3548	0.2050	3
100	20	1000	N/A	N/A	N/A	1
100	50	50	2.2202	231.0243	0.2505	24
100	50	100	1.4622	414.2773	0.2502	7
100	50	500	0.9920	722.3548	0.2050	3
100	50	1000	N/A	N/A	N/A	1
100	1280	50	1.4979	253.2288	0.2113	12
100	1280	100	1.2666	423.9990	0.2332	7
100	1280	500	0.9920	722.3548	0.2050	3
100	1280	1000	N/A	N/A	N/A	1

Table A.8: OPTICS results using multimodal embeddings.

Neighbors	Components	Min samples	Davies-Bouldin index	Calinski-Harabasz Index	Silhouette Coefficient	Number of clusters
5	3	50	3.185	88.534	0.180	30
5	3	100	2.683	195.801	0.151	10
5	3	500	2.341	767.200	0.177	3
5	3	1000	N/A	N/A	N/A	1
5	5	50	2.093	222.723	0.168	16
5	5	100	2.335	300.999	-0.078	8
5	5	500	2.385	644.456	0.147	3
5	5	1000	N/A	N/A	N/A	1
5	20	50	1.955	212.145	0.126	15
5	20	100	2.416	377.623	0.212	8
5	20	500	2.423	669.377	0.160	3
5	20	1000	N/A	N/A	N/A	1
5	50	50	2.092	199.784	0.144	17
5	50	100	2.192	415.183	0.181	6
5	50	500	2.462	625.823	0.153	3
5	50	1000	N/A	N/A	N/A	1
5	1280	50	2.133	95.435	0.016	23
5	1280	100	2.316	383.590	0.160	6
5	1280	500	2.244	638.770	0.146	3
5	1280	1000	N/A	N/A	N/A	1
50	3	50	2.202	197.120	0.222	26
50	3	100	2.157	428.877	0.251	7
50	3	500	0.996	722.258	0.205	3
50	3	1000	N/A	N/A	N/A	1
50	5	50	2.225	207.562	0.247	27
50	5	100	2.144	400.770	0.247	8
50	5	500	1.645	726.461	0.202	3
50	5	1000	N/A	N/A	N/A	1
50	20	50	2.227	215.633	0.240	24
50	20	100	1.159	418.769	0.249	7
50	20	500	0.996	722.258	0.205	3
50	20	1000	N/A	N/A	N/A	1
50	50	50	1.753	245.030	0.214	14
50	50	100	1.158	418.609	0.249	7
50	50	500	1.401	708.527	0.202	3
50	50	1000	N/A	N/A	N/A	1
50	1280	50	1.475	249.530	0.207	12
50	1280	100	1.158	418.652	0.249	7
50	1280	500	0.997	722.244	0.205	3
50	1280	1000	N/A	N/A	N/A	1
100	3	50	2.207	208.571	0.228	25
100	3	100	1.565	439.998	0.230	6
100	3	500	0.991	722.369	0.205	3
100	3	1000	N/A	N/A	N/A	1
100	5	50	1.435	253.126	0.210	12
100	5	100	1.509	408.994	0.247	7
100	5	500	0.991	722.369	0.205	3

100	5	1000	N/A	N/A	N/A	1
100	20	50	1.694	248.691	0.206	12
100	20	100	1.317	461.943	0.248	6
100	20	500	0.992	722.355	0.205	3
100	20	1000	N/A	N/A	N/A	1
100	50	50	2.135	204.211	0.221	25
100	50	100	1.319	459.964	0.247	6
100	50	500	0.992	722.355	0.205	3
100	50	1000	N/A	N/A	N/A	1
100	1280	50	1.651	250.758	0.206	12
100	1280	100	1.202	564.065	0.246	5
100	1280	500	0.991	722.369	0.205	3
100	1280	1000	N/A	N/A	N/A	1

Table A.9: HDBSCAN results using multimodal embeddings.