INM - EXPLAIN

J. Duckes, A. Girondin, M. Maïo, H. Sayah, **A. Zouzou**, S. Benslimane, J. Azé, S. Bringay, C. Mollevi, M. Servajean

INTRODUCTION

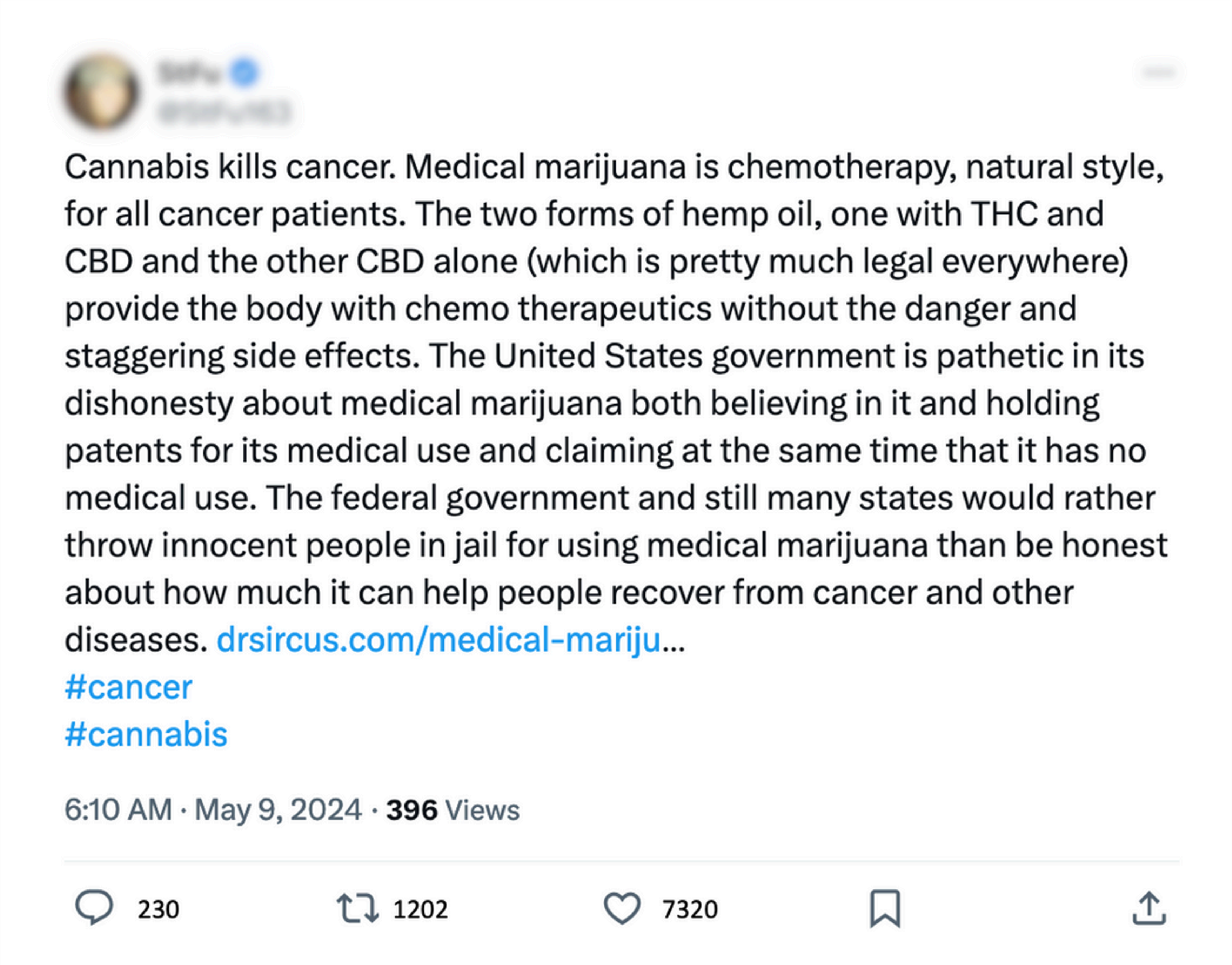
- Academic Project - Master MIASHS
- Continuation of S.Benslimane’s thesis work on the controversy - Deep Learning and Graphs to produce a controversy score

- **Our mission :** To create simple visualizations to understand the controversy
- **Case study :** INM (Non-Pharmaceutical Interventions), a highly controversial topic

---

1. **Samy Benslimane, Thèse de doctorat de l'Université de Montpellier soutenance le 08/12/2023, Analyse, détection et quantification de la controverse dans les médias sociaux par l’utilisation de modèle d’apprentissage profond sur des données structurelles et textuelles**
2. **Centre National de Ressources Textuelles et Lexicales. (n.d.). Controverse. In CNRTL.**

INTRODUCTION



Cannabis kills cancer. Medical marijuana is chemotherapy, natural style, for all cancer patients. The two forms of hemp oil, one with THC and CBD and the other CBD alone (which is pretty much legal everywhere) provide the body with chemo therapeutics without the danger and staggering side effects. The United States government is pathetic in its dishonesty about medical marijuana both believing in it and holding patents for its medical use and claiming at the same time that it has no medical use. The federal government and still many states would rather throw innocent people in jail for using medical marijuana than be honest about how much it can help people recover from cancer and other diseases. drsircus.com/medical-mariju...
#cancer
#cannabis

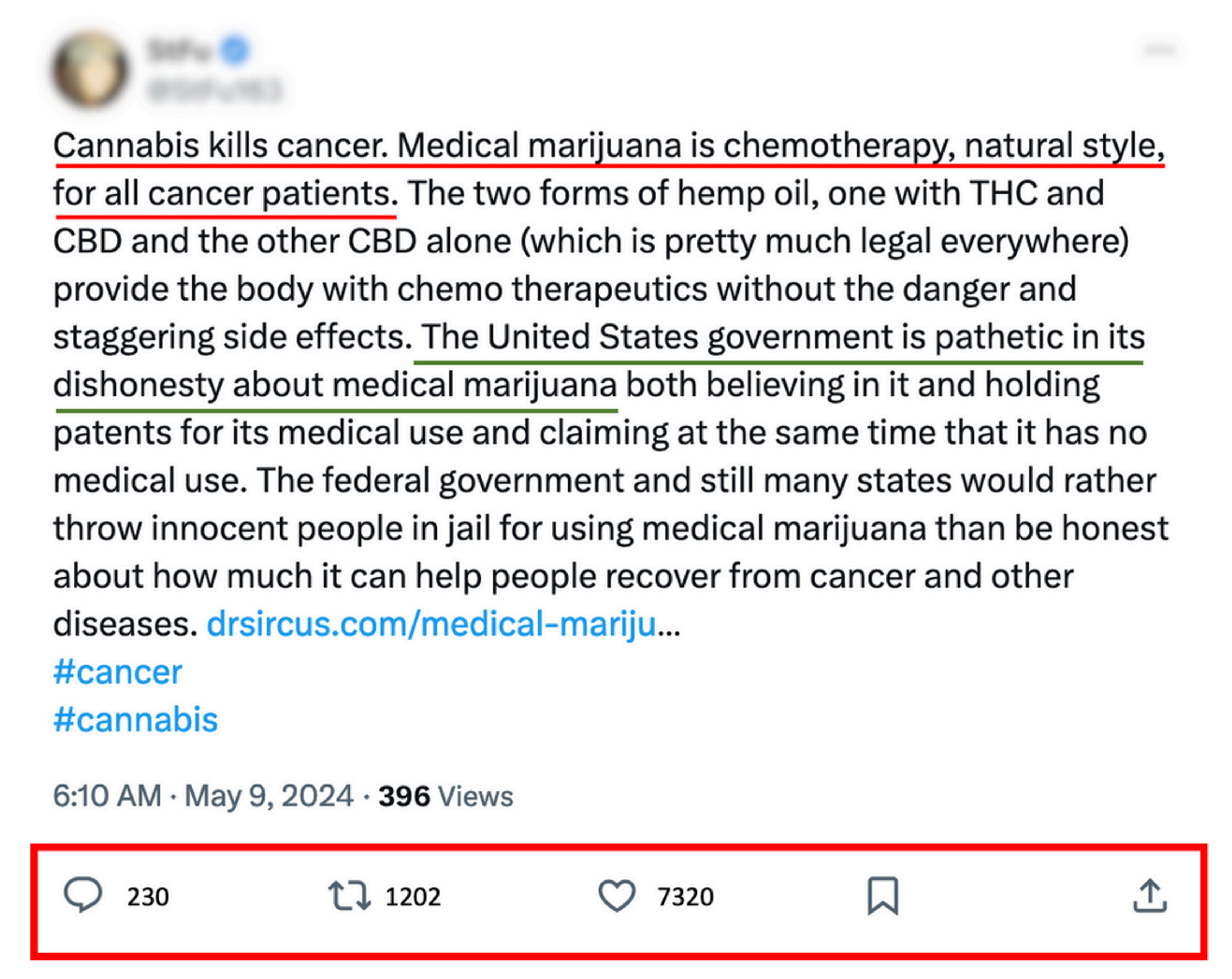
6:10 AM · May 9, 2024 · **396** Views

- **The Controversy :** A reasoned discussion, a dispute over an opinion, [...] encompassing the divergent or contradictory elements of the debate. (National Center for Textual and Lexical Resources)
- **Non-Medical Interventions (INM):** Non-invasive and non-pharmacological interventions on human health based on science. (CEPS Platform)

---

1. Samy Benslimane, Thèse de doctorat de l'Université de Montpellier soutenance le 08/12/2023, Analyse, détection et quantification de la controverse dans les médias sociaux par l'utilisation de modèle d'apprentissage profond sur des données structurelles et textuelles
2. Centre National de Ressources Textuelles et Lexicales. (n.d.). Controverse. In CNRTL.

INTRODUCTION



Cannabis kills cancer. Medical marijuana is chemotherapy, natural style, for all cancer patients. The two forms of hemp oil, one with THC and CBD and the other CBD alone (which is pretty much legal everywhere) provide the body with chemo therapeutics without the danger and staggering side effects. The United States government is pathetic in its dishonesty about medical marijuana both believing in it and holding patents for its medical use and claiming at the same time that it has no medical use. The federal government and still many states would rather throw innocent people in jail for using medical marijuana than be honest about how much it can help people recover from cancer and other diseases. drsircus.com/medical-mariju...
#cancer
#cannabis

6:10 AM · May 9, 2024 · **396** Views

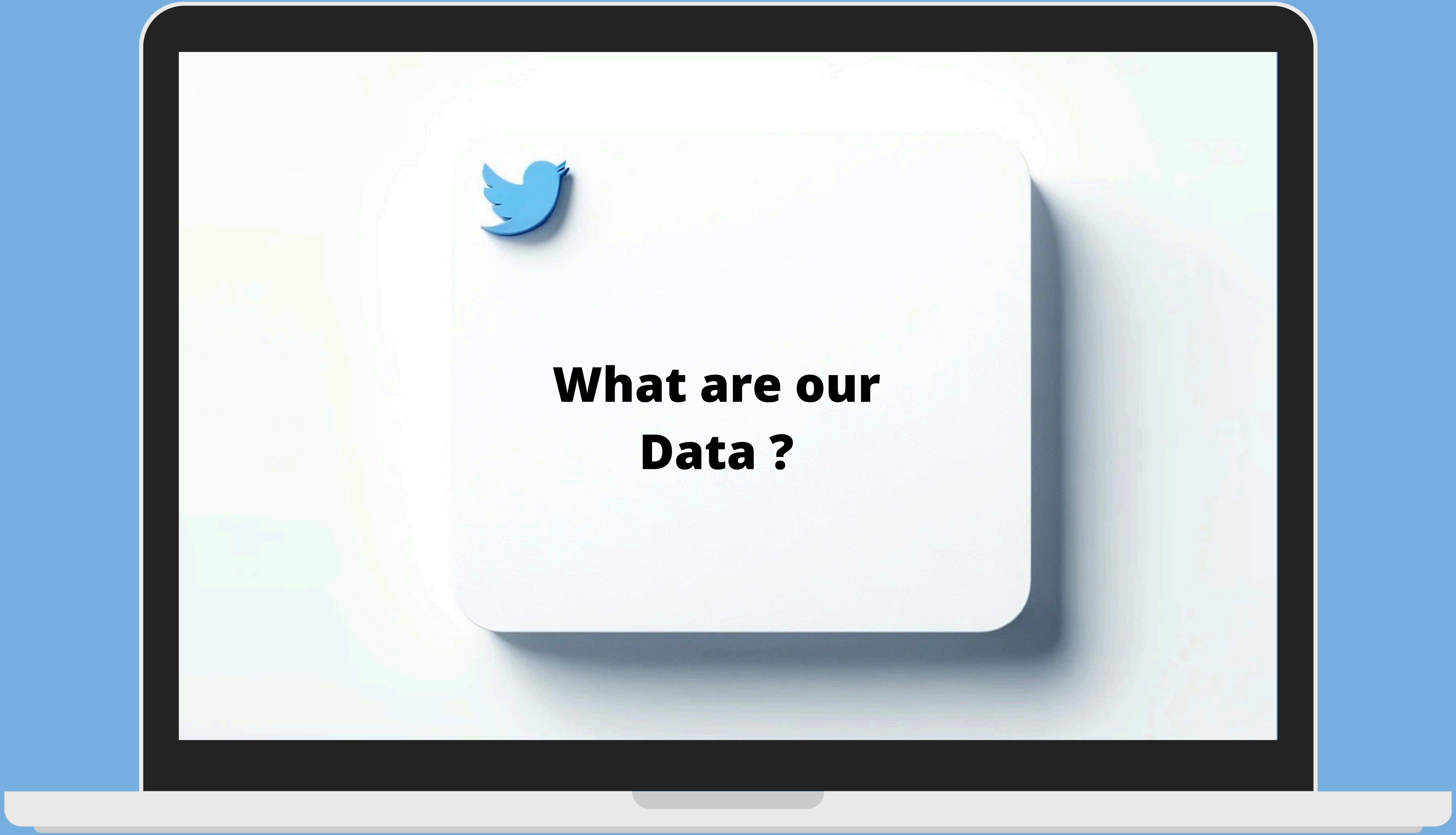
How can data visualizations contribute to a better understanding of the factors behind controversies on Twitter?

---

1. Samy Benslimane, Thèse de doctorat de l'Université de Montpellier soutenance le 08/12/2023, Analyse, détection et quantification de la controverse dans les médias sociaux par l'utilisation de modèle d'apprentissage profond sur des données structurelles et textuelles
2. Centre National de Ressources Textuelles et Lexicales. (n.d.). Controverse. In CNRTL.

#Summary

- Data Processing 2
- Developed Visualizations 3
- Where to find our results ? 11
- Discussions and Perspectives 12
- Conclusion 13



What are our Data ?

Data Processing

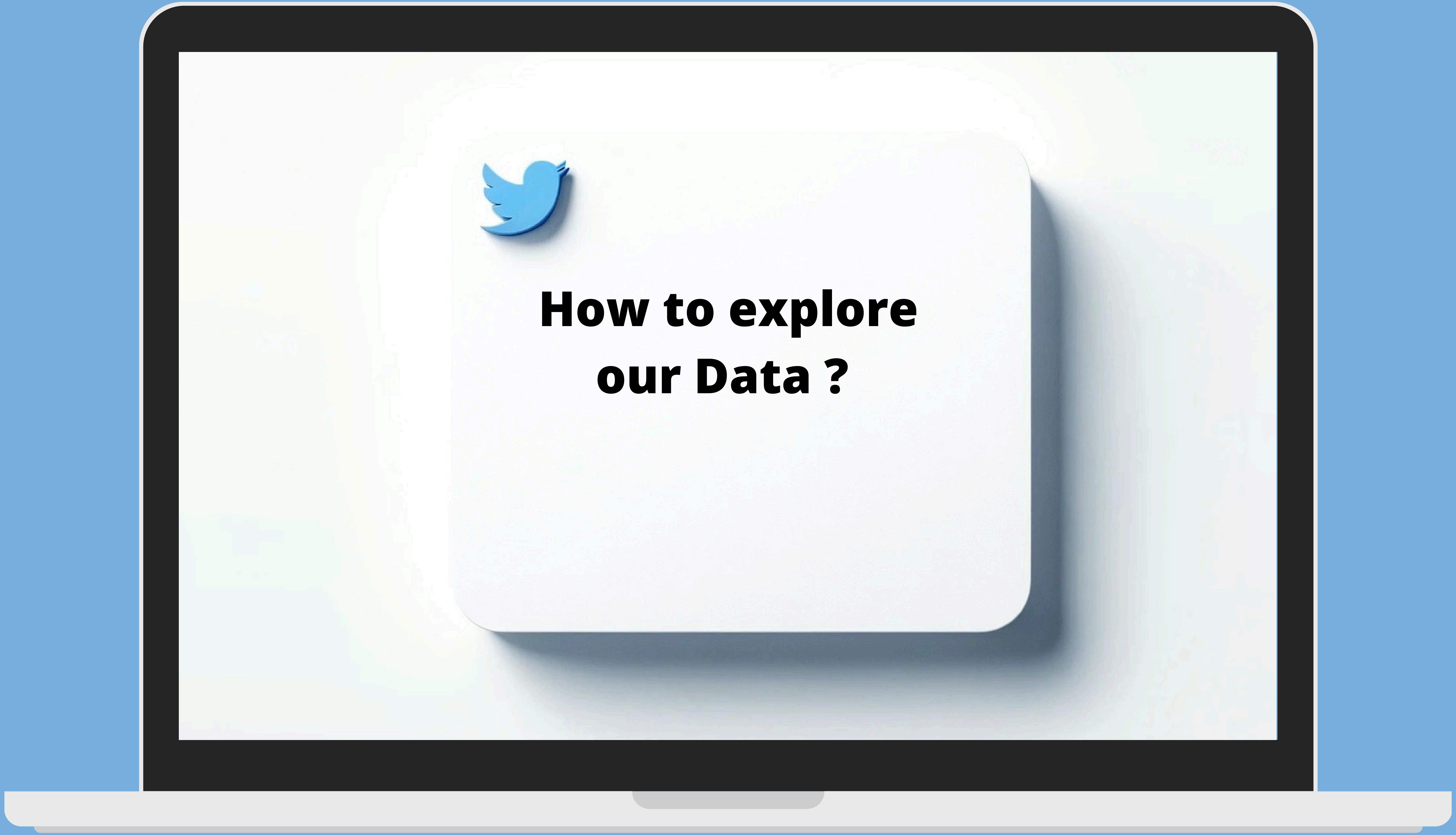
Tweets : 52 920
Likes : 7 209 886
Retweets : 2 048 663
Responses : 623 221
Citations : 280 448

INM

Data Processing

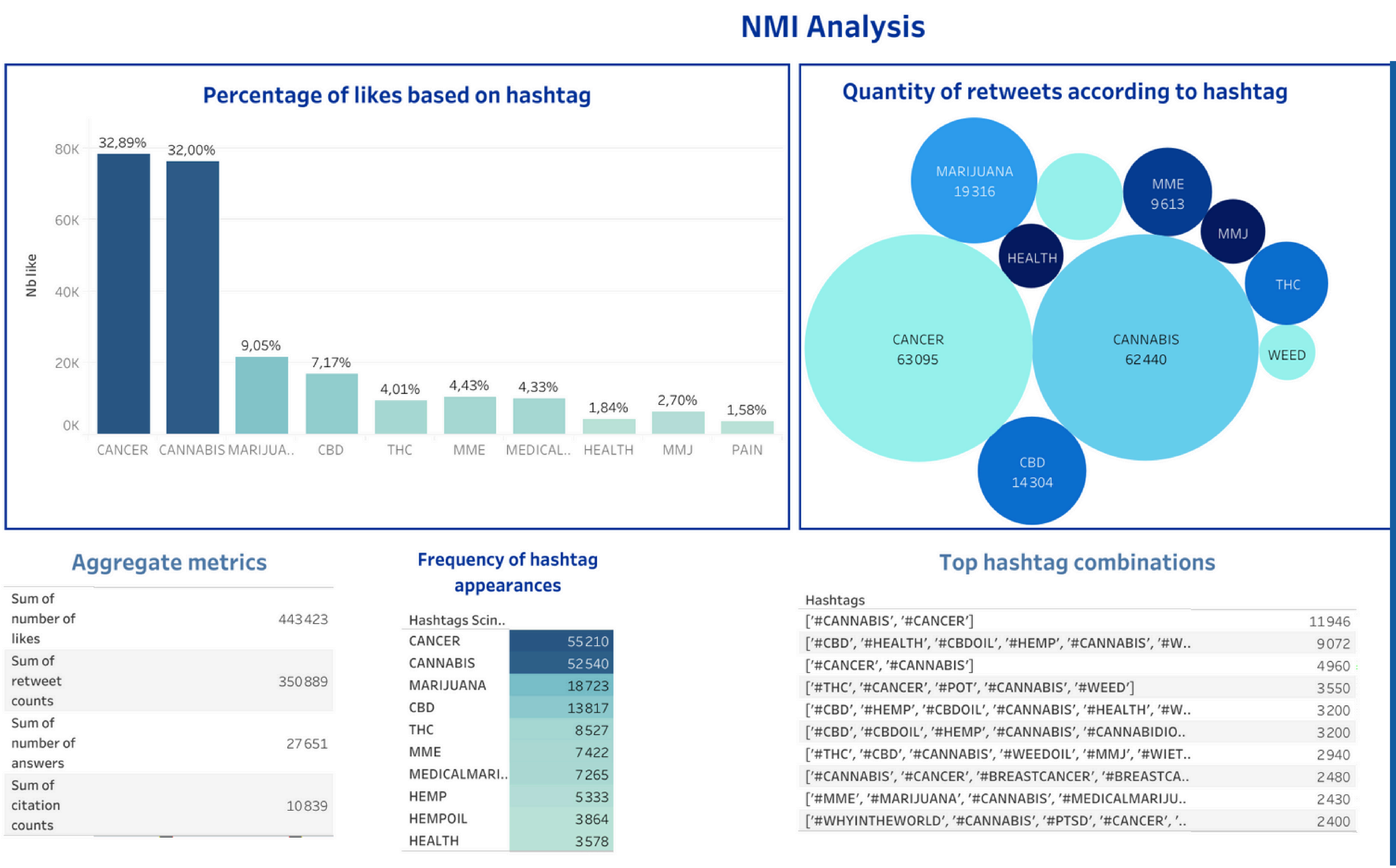
- Tokenization
- Hashtags Retrieval
- Users Removal

- LDA
- Clustering and Community Detection
- Sentiment Analysis
- Opinion Classification
- Contreversy index



How to explore our Data ?

Initial Explorations



NMI Analysis

Percentage of likes based on hashtag

Nb like
80K
60K
40K
20K
0K
32,89%
32,00%
9,05%
7,17%
4,01%
4,43%
4,33%
1,84%
2,70%
1,58%
CANCER
CANNABIS
MARIJUA..
CBD
THC
MME
MEDICAL..
HEALTH
MMJ
PAIN

Quantity of retweets according to hashtag

MARIJUANA
19 316
MME
9 613
MMJ
HEALTH
THC
CANCER
63 095
CANNABIS
62 440
WEED
CBD
14 304

Aggregate metrics

Sum of number of likes	443 423
Sum of retweet counts	350 889
Sum of number of answers	27 651
Sum of citation counts	10 839

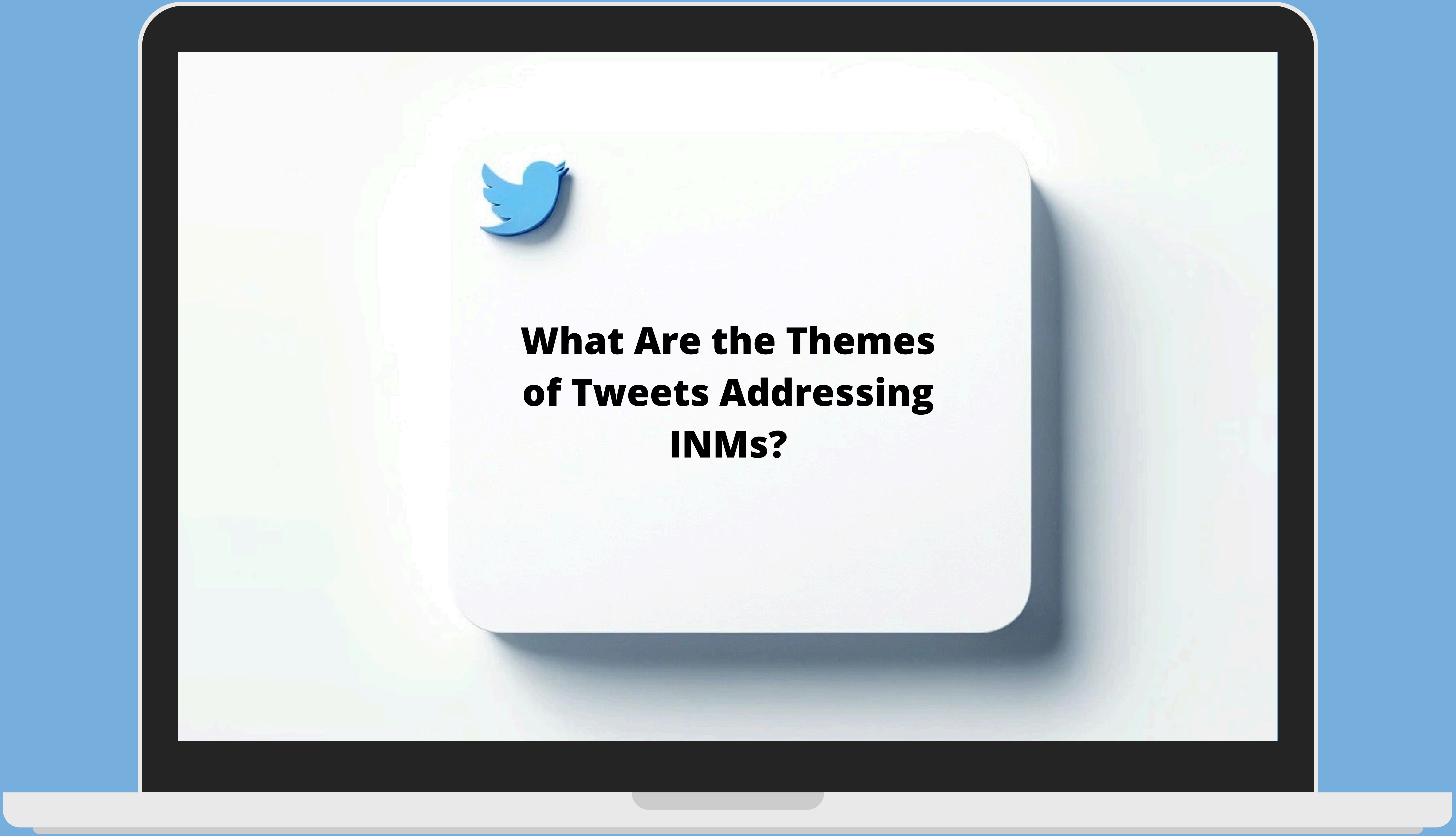
Frequency of hashtag appearances

Hashtags Scin..	
CANCER	55 210
CANNABIS	52 540
MARIJUANA	18 723
CBD	13 817
THC	8 527
MME	7 422
MEDICALMARI..	7 265
HEMP	5 333
HEMPOIL	3 864
HEALTH	3 578

Top hashtag combinations

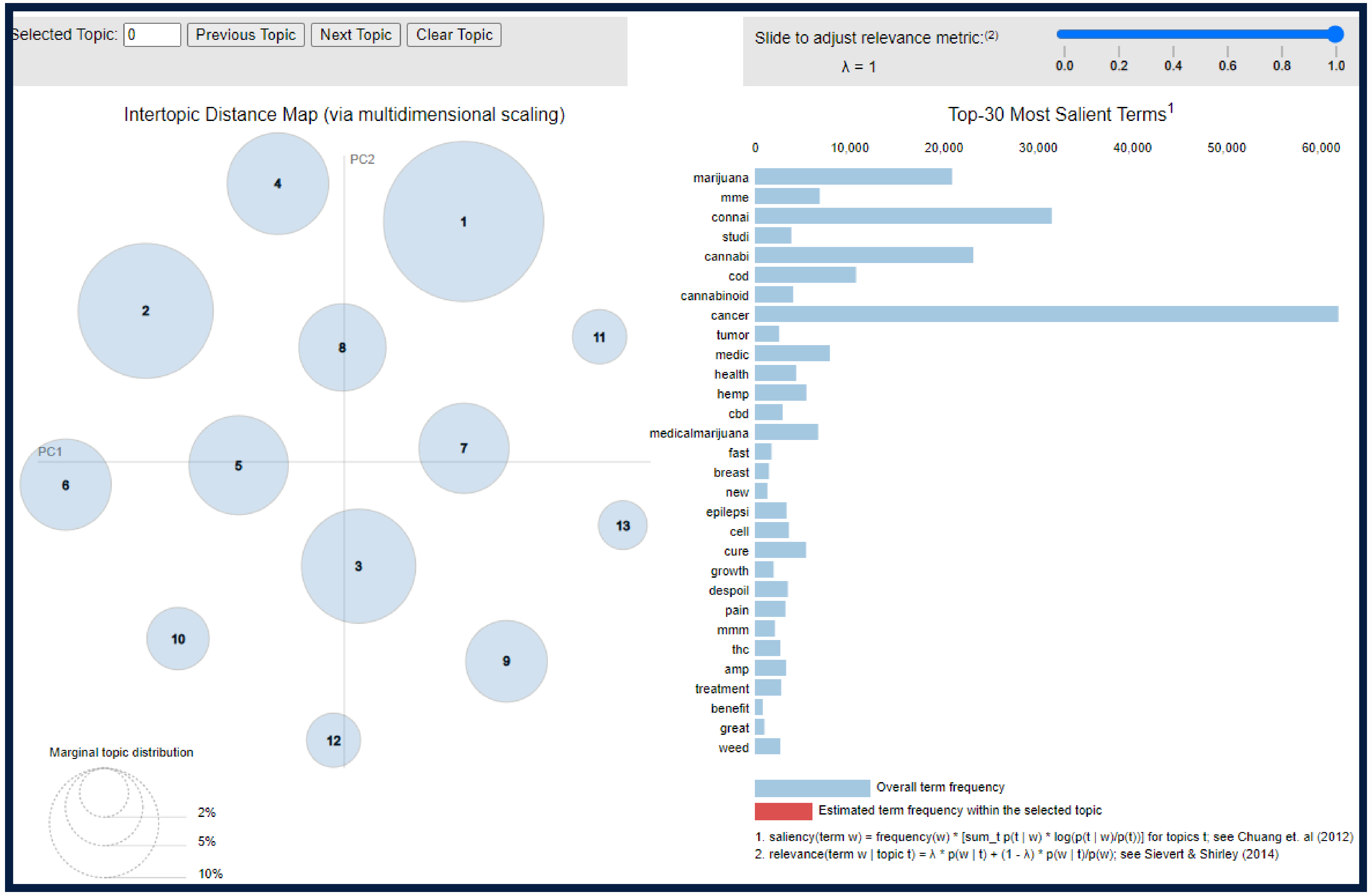
Hashtags	
['#CANNABIS', '#CANCER']	11 946
['#CBD', '#HEALTH', '#CBDOIL', '#HEMP', '#CANNABIS', '#W..	9 072
['#CANCER', '#CANNABIS']	4 960
['#THC', '#CANCER', '#POT', '#CANNABIS', '#WEED']	3 550
['#CBD', '#HEMP', '#CBDOIL', '#CANNABIS', '#HEALTH', '#W..	3 200
['#CBD', '#CBDOIL', '#HEMP', '#CANNABIS', '#CANNABIDIO..	3 200
['#THC', '#CBD', '#CANNABIS', '#WEEDOIL', '#MMJ', '#WIET..	2 940
['#CANNABIS', '#CANCER', '#BREASTCANCER', '#BREASTCA..	2 480
['#MME', '#MARIJUANA', '#CANNABIS', '#MEDICALMARIJU..	2 430
['#WHYINTHEWORLD', '#CANNABIS', '#PTSD', '#CANCER', '..	2 400

- visualizations on **Tableau**
- **#CANCER #CANNABIS** are "top hashtags"



What Are the Themes of Tweets Addressing INMs?

Latent Dirichlet Allocation - Topic Modeling


Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
6
7
8
9
10
11
12
13
Marginal topic distribution
2%
5%
10%
Top-30 Most Salient Terms1
0
10,000
20,000
30,000
40,000
50,000
60,000
marijuana
mme
connai
studi
cannabi
cod
cannabinoid
cancer
tumor
medic
health
hemp
cbd
medicalmarijuana
fast
breast
new
epilepsi
cell
cure
growth
despoil
pain
mmm
thc
amp
treatment
benefit
great
weed
Overall term frequency
Estimated term frequency within the selected topic

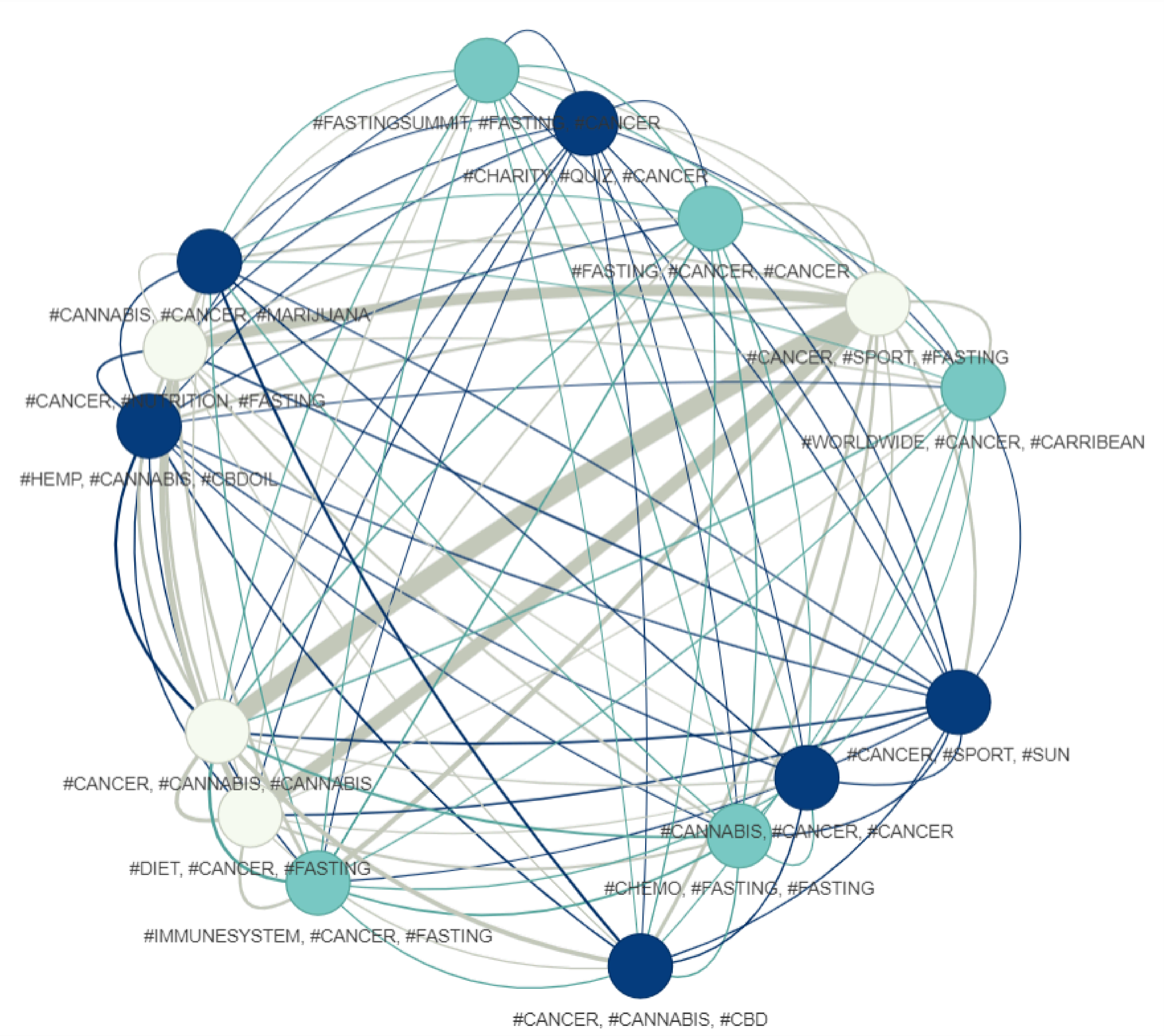
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Label	Top Words
Treatment with Cannabis	[cancer, cannabi, connai, cell, marijuana, cure, studi, treatment, cannabinoid, kill]
Diet and Intermittent Fasting	[cancer, cannabi, breast, connai, breastcanc, analysi, extens, v, sport, wast]
Health Benefits of Cannabis	[cancer, health, daili, connai, cannabi, latest, benefit, cbd, hemp, ill]
Medical Marijuana Usage	[cancer, connai, cannabi, marijuana, leukaemia, medicalmarijuana, mmm, health, medicine, cod]
Research on Cancer Prevention	[cancer, connai, cannabi, amp, pain, fight, effect, growth, patient, cell]
Health and CBD Oil Usage	[cancer, connai, cb, cod, hemp, list, mmm, e, wietoli, weedoil]

Chaney & Blei. Visualizing topic models. In Proceedings of the International AAAI Conference on Web and Social Media (2012 Vol. 6, No. 1, pp. 419-422).

How Can We Explain the Interactions Between Users Within Communities?

Community Detection Within Hashtag Clusters


#FASTINGSUMMIT, #FASTING, #CANCER
#CHARITY, #QUIZ, #CANCER
#FASTING, #CANCER, #CANCER
#CANNABIS, #CANCER, #MARIJUANA
#CANCER, #SPORT, #FASTING
#CANCER, #NUTRITION, #FASTING
#WORLDWIDE, #CANCER, #CARRIBEAN
#HEMP, #CANNABIS, #CBDOIL
#CANCER, #SPORT, #SUN
#CANCER, #CANNABIS, #CANNABIS
#CANNABIS, #CANCER, #CANCER
#DIET, #CANCER, #FASTING
#CHEMO, #FASTING, #FASTING
#IMMUNESYSTEM, #CANCER, #FASTING
#CANCER, #CANNABIS, #CBD

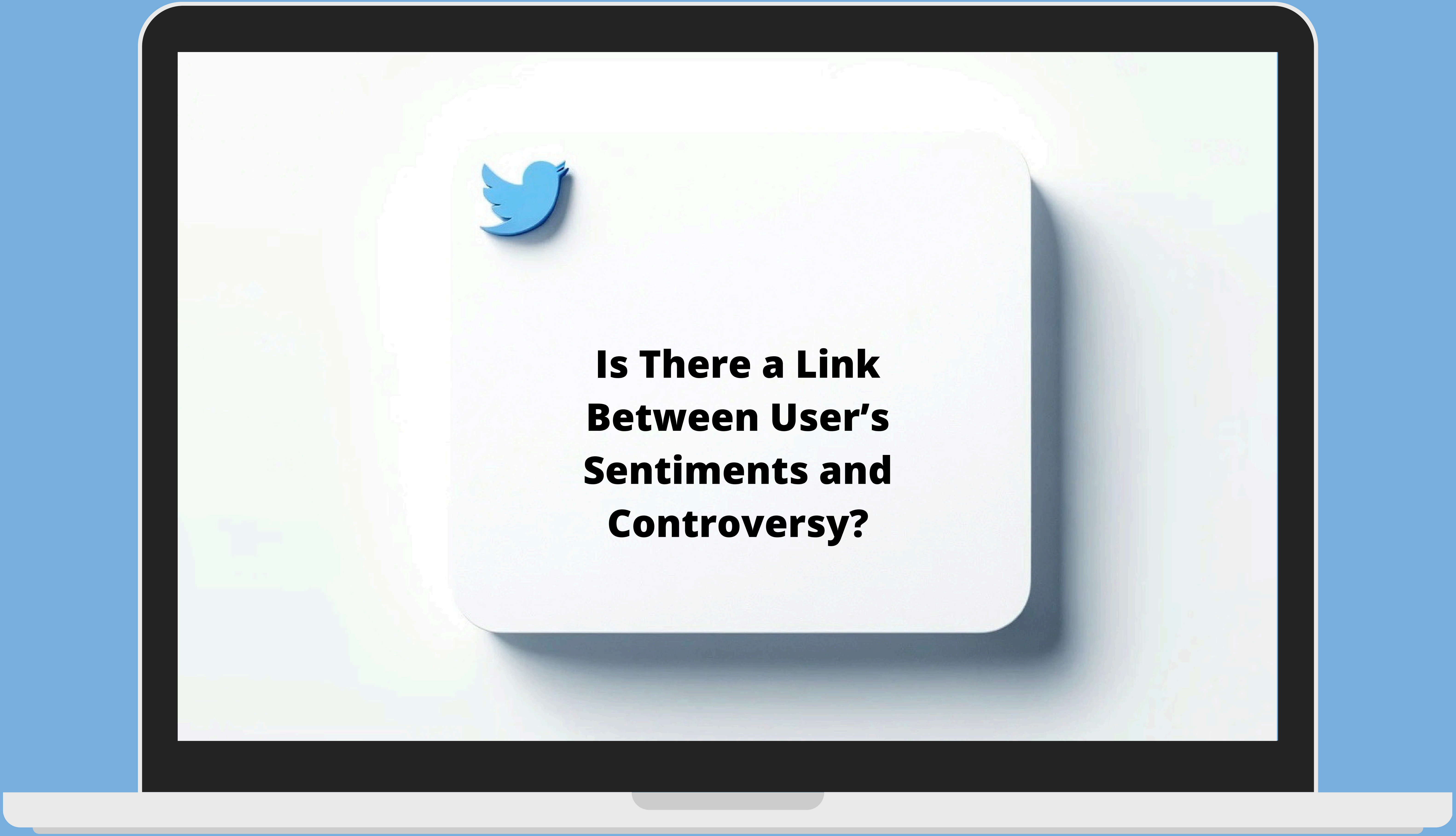
- **Links Between Hashtags in Tweets**
- **2 391 tweets** represented (1.5%)

- a node = a cluster = a similar group
- **K-means Clustering**

- Color represents community
- **Louvain's Algorithm**

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PLoS ONE, 9(6), e98679.

Que et al. Scalable Community Detection with the Louvain Algorithm, In Proceedings of the2015 IEEE International Parallel and Distributed Processing Symposium, Hyderabad, India, 2015, 28-37

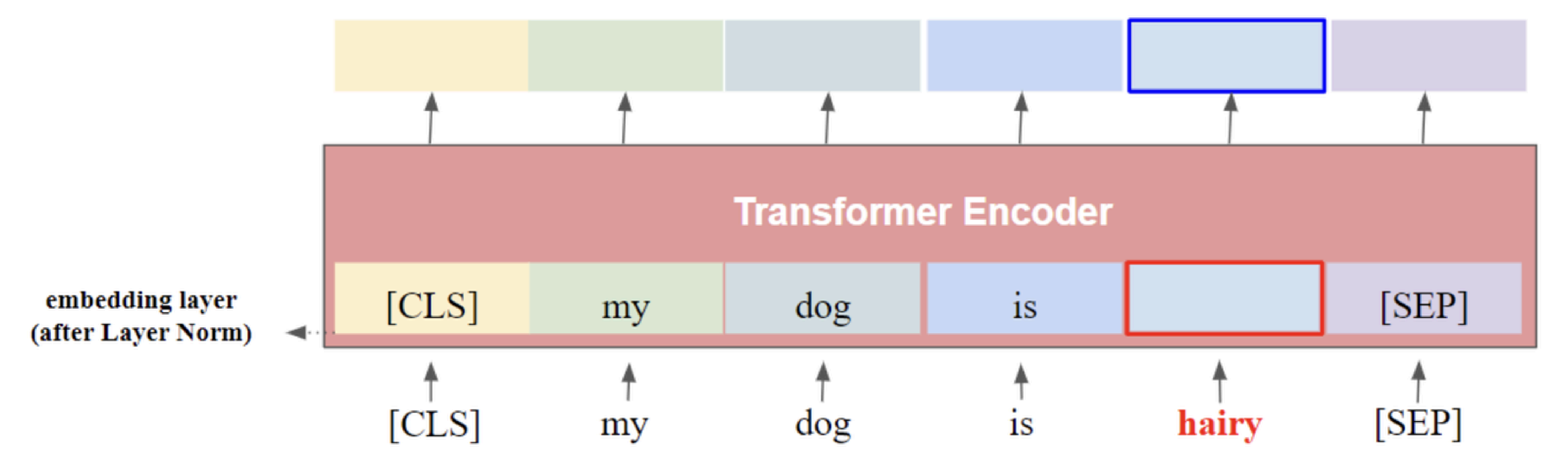


Is There a Link Between User’s Sentiments and Controversy?

Sentiment Analysis with Hugging Face

RoBERTa

- Bidirectional Understanding
- Masked Language Modeling (MLM)
- Next Sentence Prediction (NSP)


Transformer Encoder
embedding layer
(after Layer Norm)
[CLS]
my
dog
is
[SEP]
[CLS]
my
dog
is
hairy
[SEP]

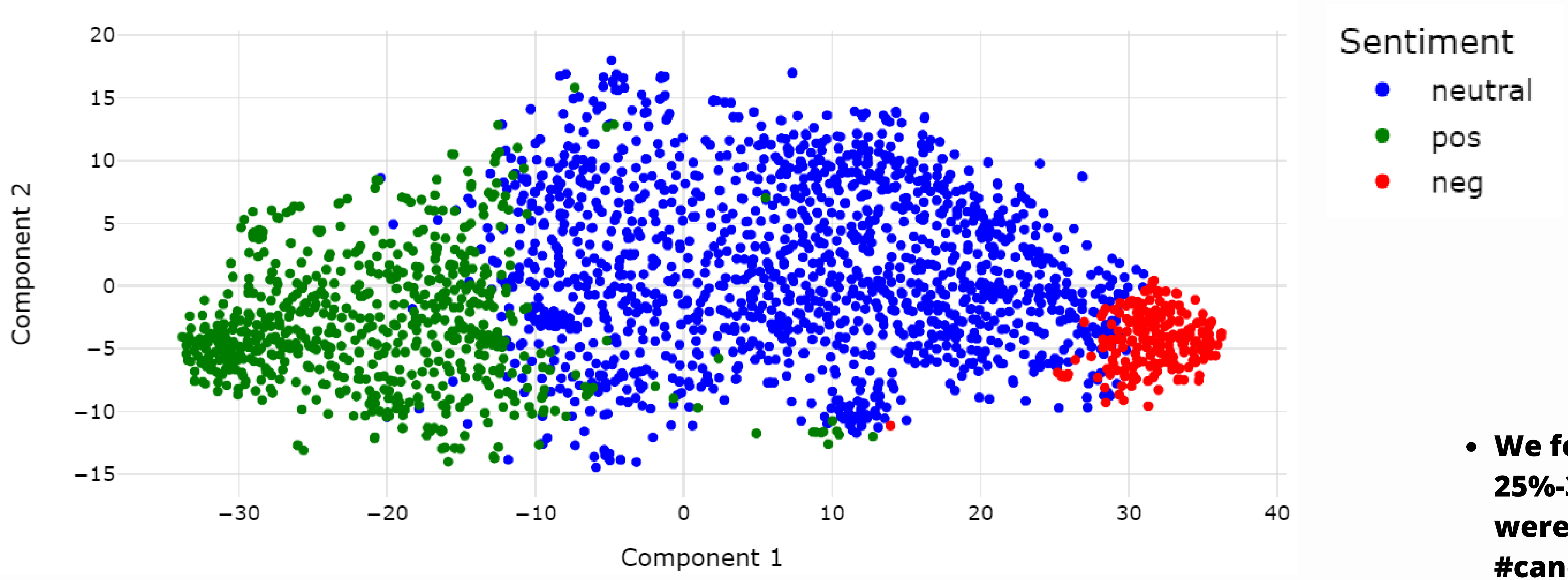
How ?

Meta

Google
Bert

Liu et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. ArXIV 2019

t-SNE Visualization of Feature Spaces for Sentiment Analysis



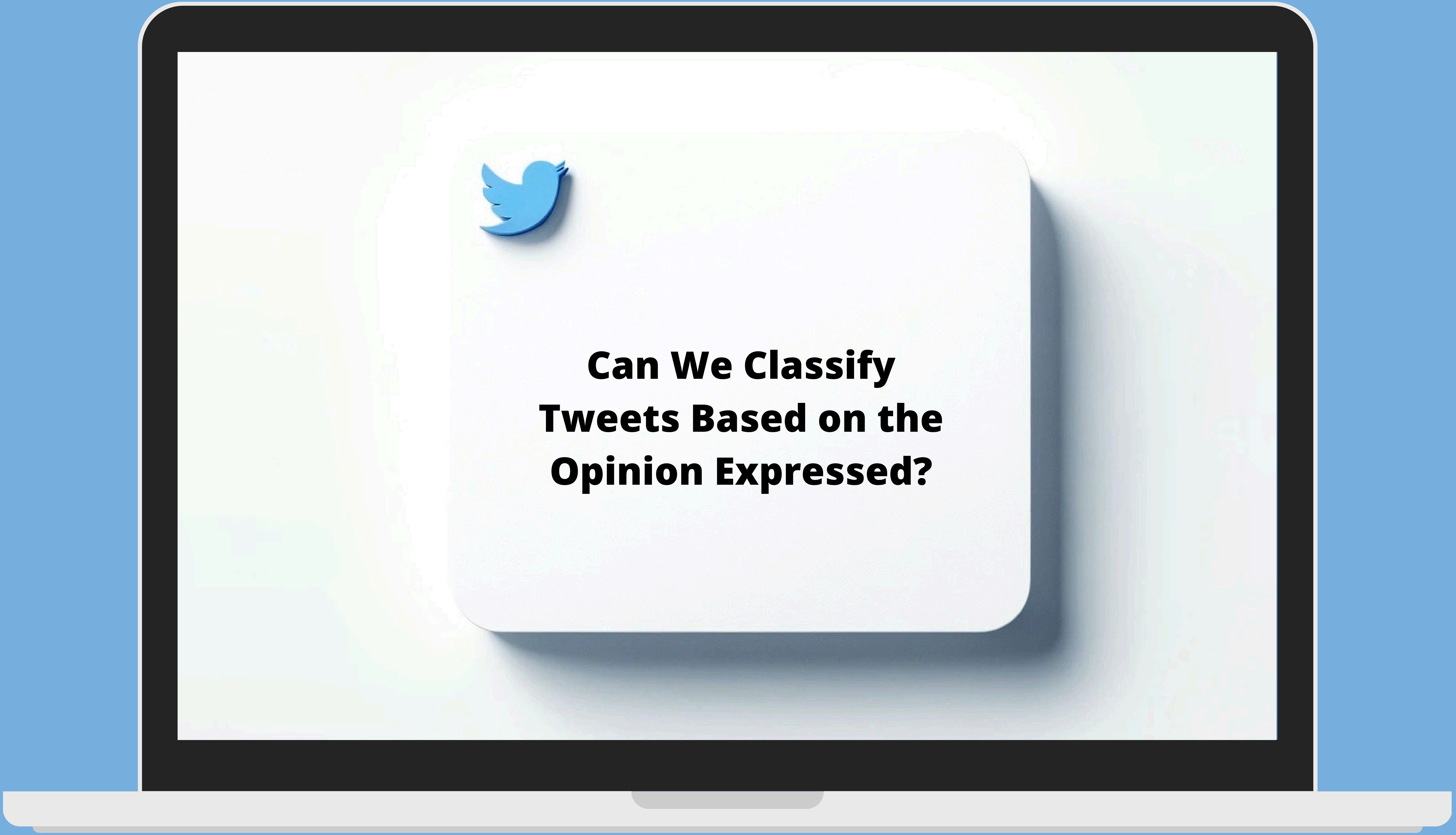
POSITIVE

- **We found that almost 25%-35% of the #mice were completely cured of #cancer after administering #Cannabis. 😄**
- **Medical Marijuana Helped Save This Child From #Cancer https://t.co/My5dO4ksKR #marijuana #cannabis #MME #Cancer http**

NEGATIVE

- **While #Cancer is big business for some, it's a business that only exists because the #Cure is illegal. #Cannabis...**
- **It doesn't matter to me if you smoke it or take drops, Just think we are healthier, saving the environment, and tax payer money, Fk big pharma. #cannabis #cbdoil #hempoil #stress #ibs #cancer #anxiety #organic #facialcream #painrub #sleep #arthritis #fibromialgia #insomnia**

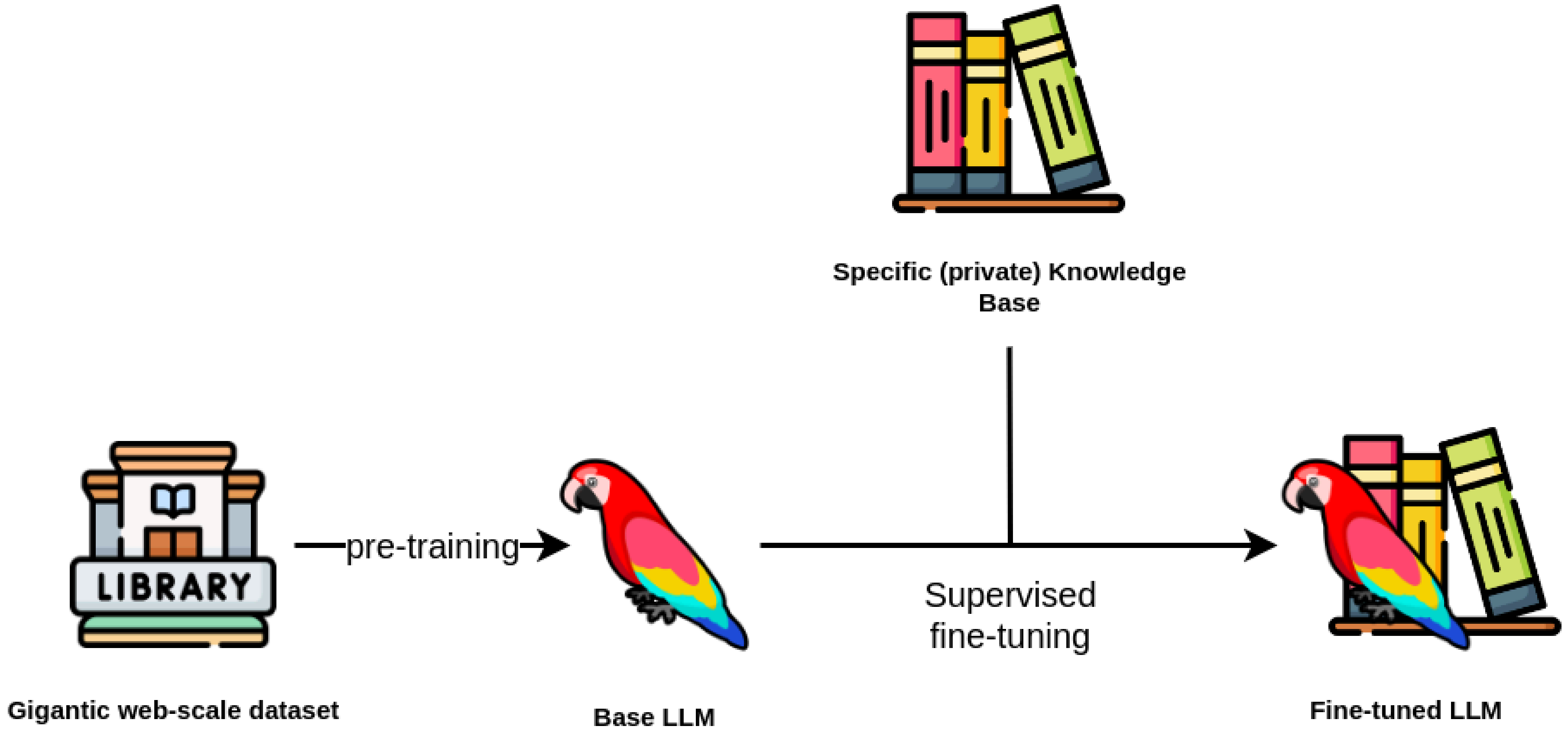
Van der Maaten & Hilton. Visualizing Data using t-SNE. Journal of Machine Learning Research 9 (2008) 2579-2605



Can We Classify Tweets Based on the Opinion Expressed?

Classification d’opinions Hugging Face

Pre-Trained model : google-bert/bert-base-uncased


Specific (private) Knowledge Base
pre-training
Supervised fine-tuning
LIBRARY
Gigantic web-scale dataset
Base LLM
Fine-tuned LLM

- Train : **80% (450 tweets)**
- Test : **20% (50 tweets)**
- **3 epochs**
- Labelisation : **PRO, ANTI, NEUTRAL**

Class Imbalance:

Minority Anti**!**

PRO : Cancer patient: 'I had to become an outlaw or die' https://t.co/iKhdQt4MDR via @user #MME #Cancer #Cannabis http

ANTI : New research suggests that #cannabis extract can have a "dramatic effect" on brain #cancer. http #breakthetaboo

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Devlin et al., NAACL 2019)

Opinion Classification

Hugging Face

Accuracy Globale : 86%

Precision	VALUE
PRO	0.971
ANTI	0.500
NEUTRAL	0.643

Problem for ANTI Class -

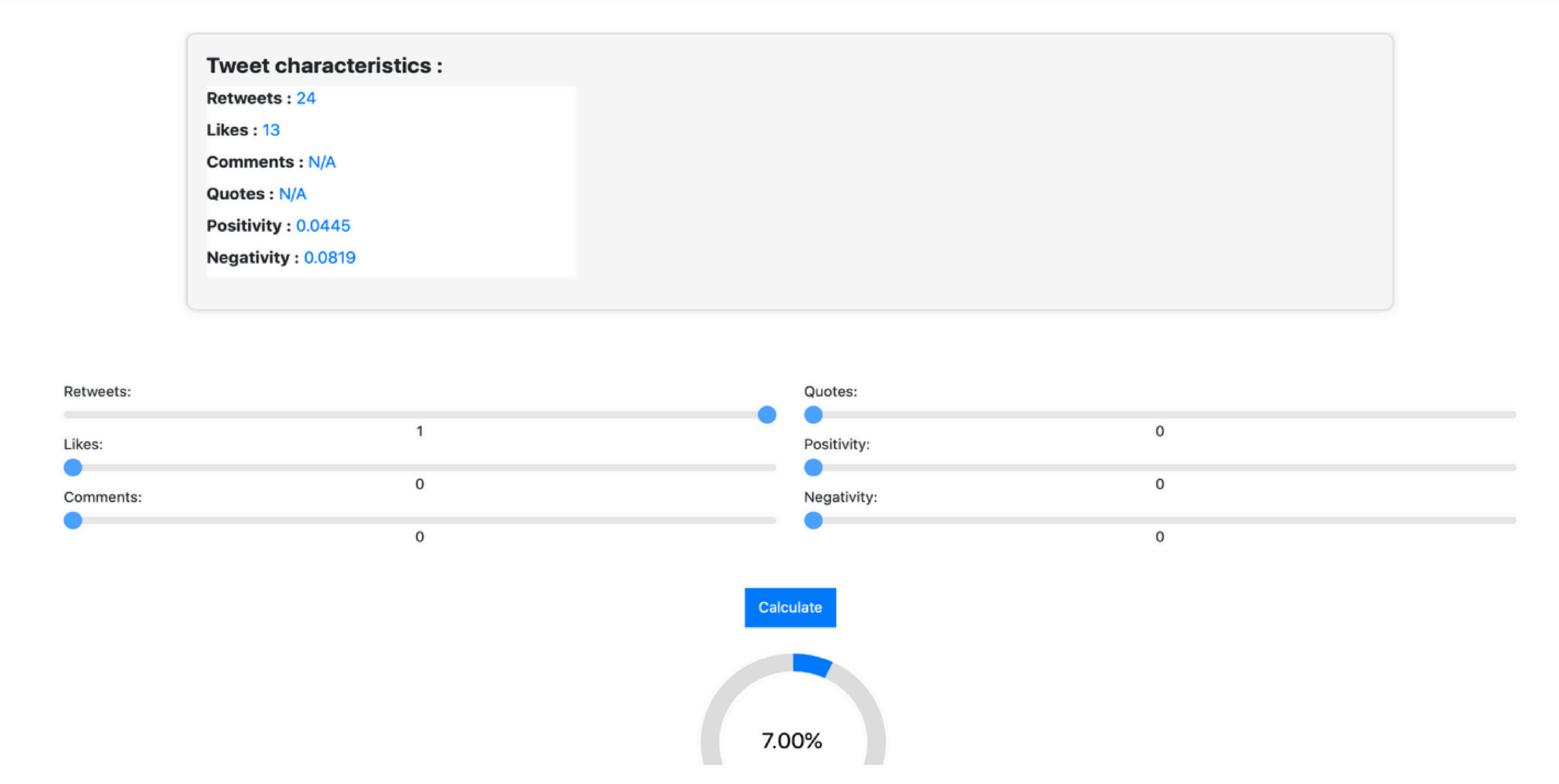
Minority

1. **DATA AUGMENTATION**
2. **LOSS to handle uneven classes**
    - Weighted Categorial Cross-Entropy Loss
    - Focal Loss

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Devlin et al., NAACL 2019)

How to quantify Contreversy?

Controversy Index



Tweet characteristics :

Retweets : 24

Likes : 13

Comments : N/A

Quotes : N/A

Positivity : 0.0445

Negativity : 0.0819

Retweets: 1

Likes: 0

Comments: 0

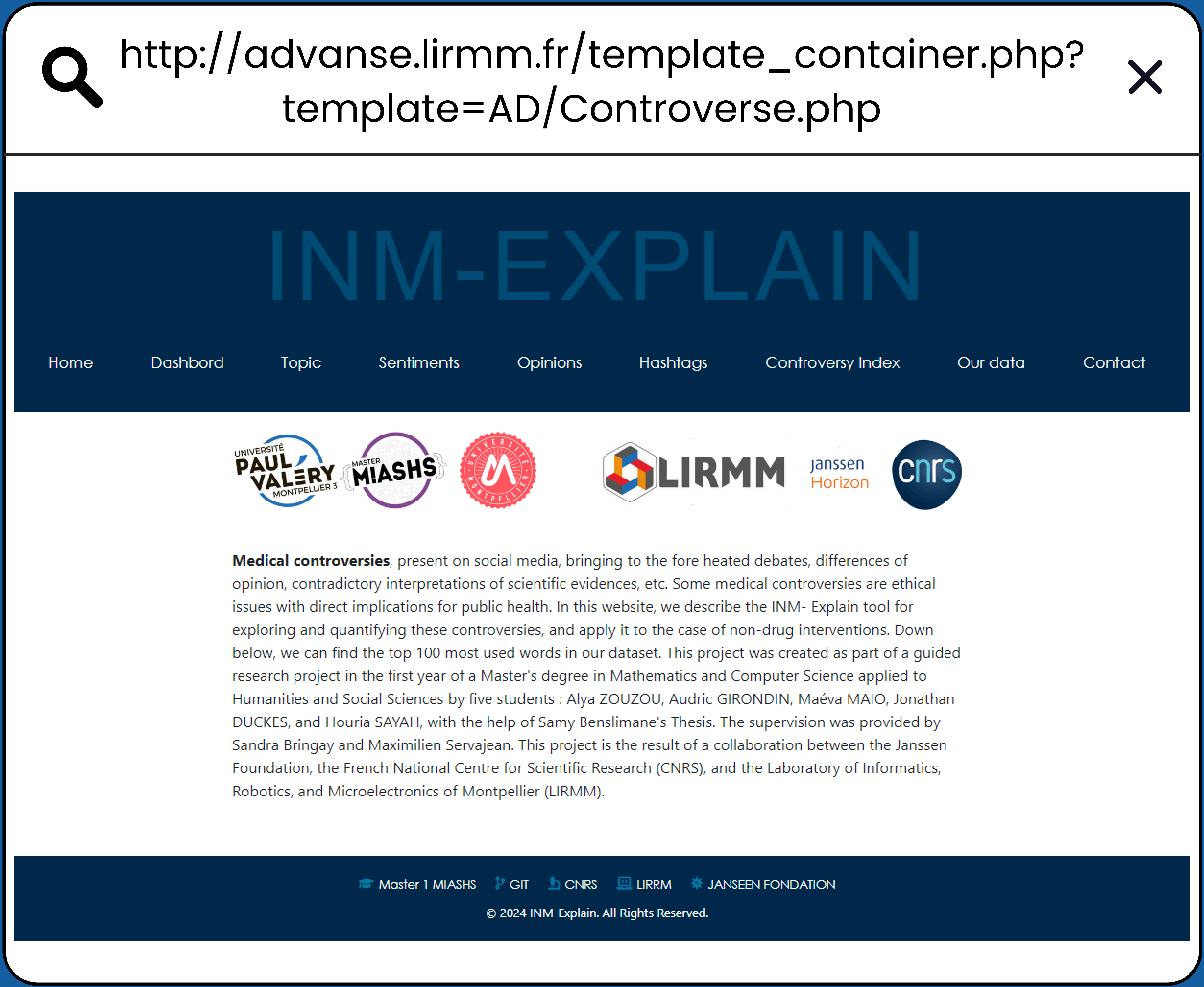
Quotes: 0

Positivity: 0

Negativity: 0

- Essential Metric
- **Multiple controversy factors are numerically normalized between 0 and 1**
- Weighted sum of factors chosen by the user.

Where to Find Our Research?



http://advanse.lirmm.fr/template_container.php?template=AD/Controverse.php

INM-EXPLAIN

Home
Dashbord
Topic
Sentiments
Opinions
Hashtags
Controversy Index
Our data
Contact

Medical controversies, present on social media, bringing to the fore heated debates, differences of opinion, contradictory interpretations of scientific evidences, etc. Some medical controversies are ethical issues with direct implications for public health. In this website, we describe the INM- Explain tool for exploring and quantifying these controversies, and apply it to the case of non-drug interventions. Down below, we can find the top 100 most used words in our dataset. This project was created as part of a guided research project in the first year of a Master's degree in Mathematics and Computer Science applied to Humanities and Social Sciences by five students : Alya ZOUZOU, Audric GIRONDIN, Maéva MAIO, Jonathan DUCKES, and Houria SAYAH, with the help of Samy Benslimane's Thesis. The supervision was provided by Sandra Bringay and Maximilien Servajean. This project is the result of a collaboration between the Janssen Foundation, the French National Centre for Scientific Research (CNRS), and the Laboratory of Informatics, Robotics, and Microelectronics of Montpellier (LIRMM).

Master 1 MIASHS
GIT
CNRS
LIRRM
JANSEEN FONDATION

© 2024 INM-Explain. All Rights Reserved.

Perspectives and Discussions

- **Improvement of Descriptive Visualizations:** Add interactions to navigate through subsets of tweets (keywords, number of retweets, etc.).
- **Refining Sentiment Analysis:** Link themes to the tweets of each group in the TSNE visualization.
- **Temporal Analysis :** Explore controversial topics over time.
- **Case Study and Perception of NPIs:** Generalize the INM-Explain tool to other controversial topics and social media platforms (Threads, TikTok).

CONCLUSION ✓

- **Cannabis :** Focus on cannabis and its derivatives, central role, and demand for legal recognition.
- **Sentiment Analysis :** Variety of perceptions, mixed reception of alternative therapies.
- **Controversy:** Need for informed discussions to guide public health policies.
- **Communication:** Importance of scientific communication and the engagement of health professionals.

MERCI !