PANicDEMIC

#coronavirus #covid19 #togetherapart #newnormal

20180745 Mai Tung Duong | 20160888 Ern Chern | 20140005 Minsoo Kang | 20170906 Assem Zhunis | 20180755 Duc Khanh Thi Nguyen

How people feel about pandemic?

Pandemic & Emotions

I **hate** when people do zero social distancing during the weekend and then post COVID-19 stats on Monday... MY N***A YOUR A PART OF THE PROBLEM!!!

My mom is a Frontliner and she got tested for the virus a few days ago. We're hoping for a negative but, it's hard not to **worry** cause i love her so much. 😭😔🙏

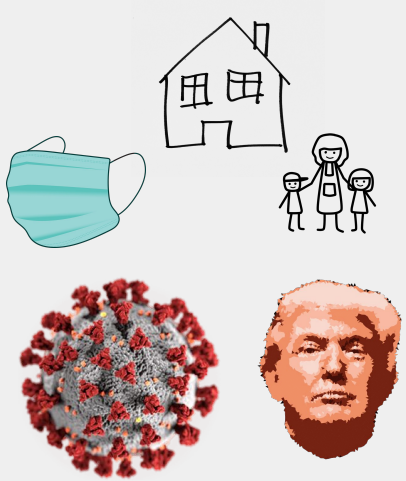
quarantine has really made me **appreciate** the little things and has shifted my focus on making myself happy rather than bending over backwards to make others **happy** :)

Covid 19 doesnt care who it takes. **Sad** that it takes many good people. **:(**

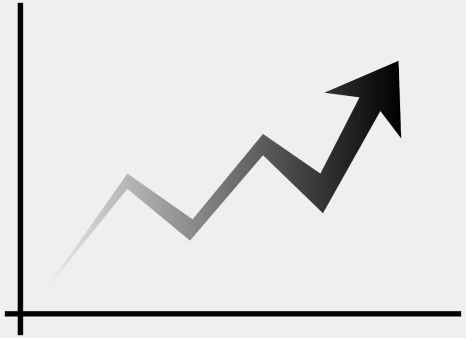
Tweets related to COVID-19

OUR GOAL

Detect emotion pattern in tweets



Find causes



Track the trend

twitter

SADNESS
INSIDE OUT

FEAR
INSIDE OUT

JOY
INSIDE OUT

ANGER
INSIDE OUT

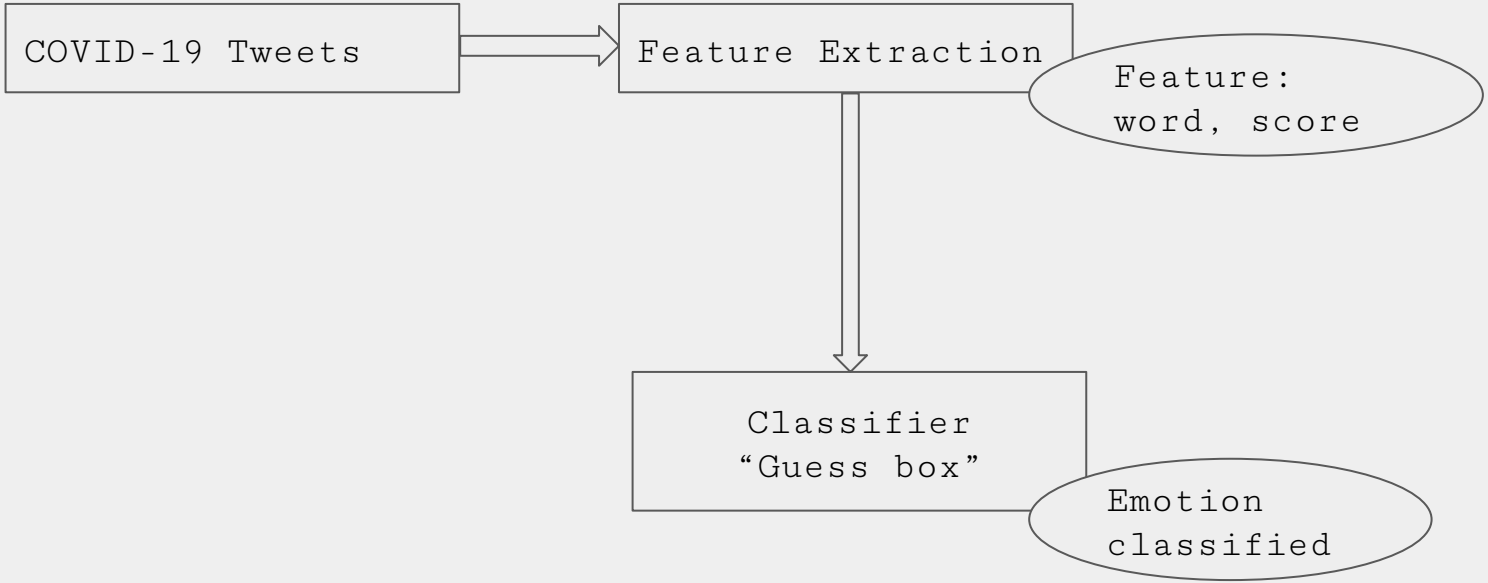
1. PIPELINE

PIPELINE (initial)

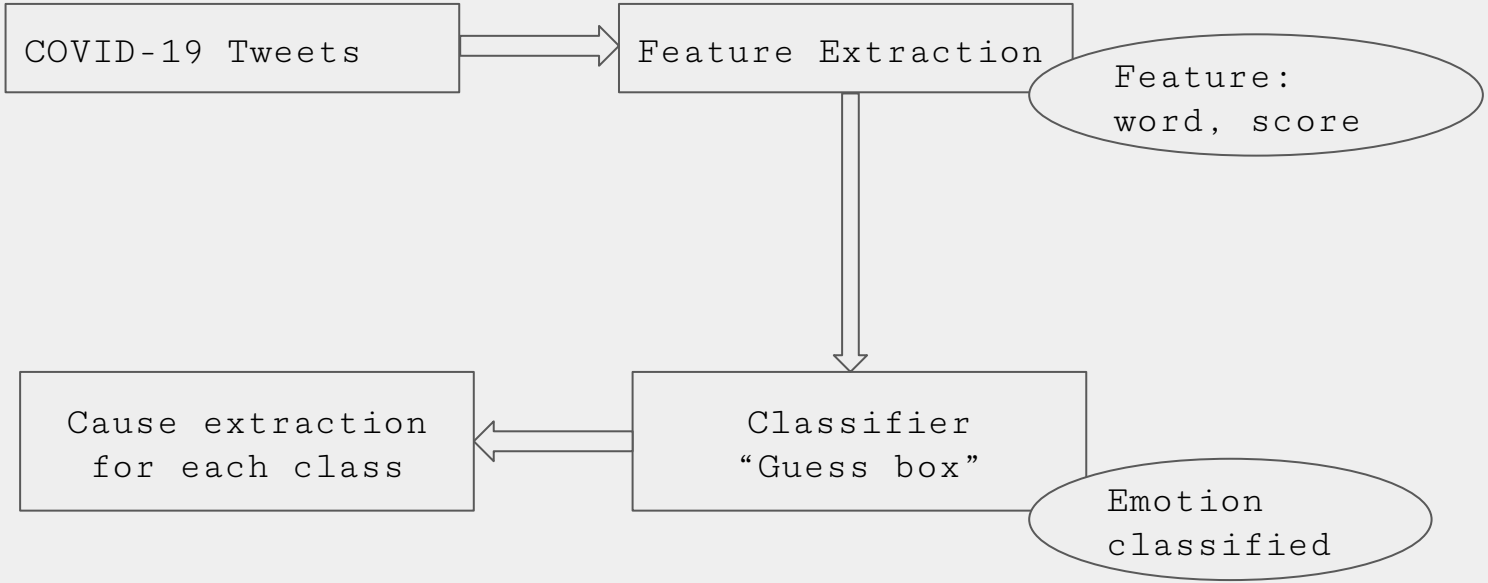
COVID-19 Tweets

PIPELINE (initial)

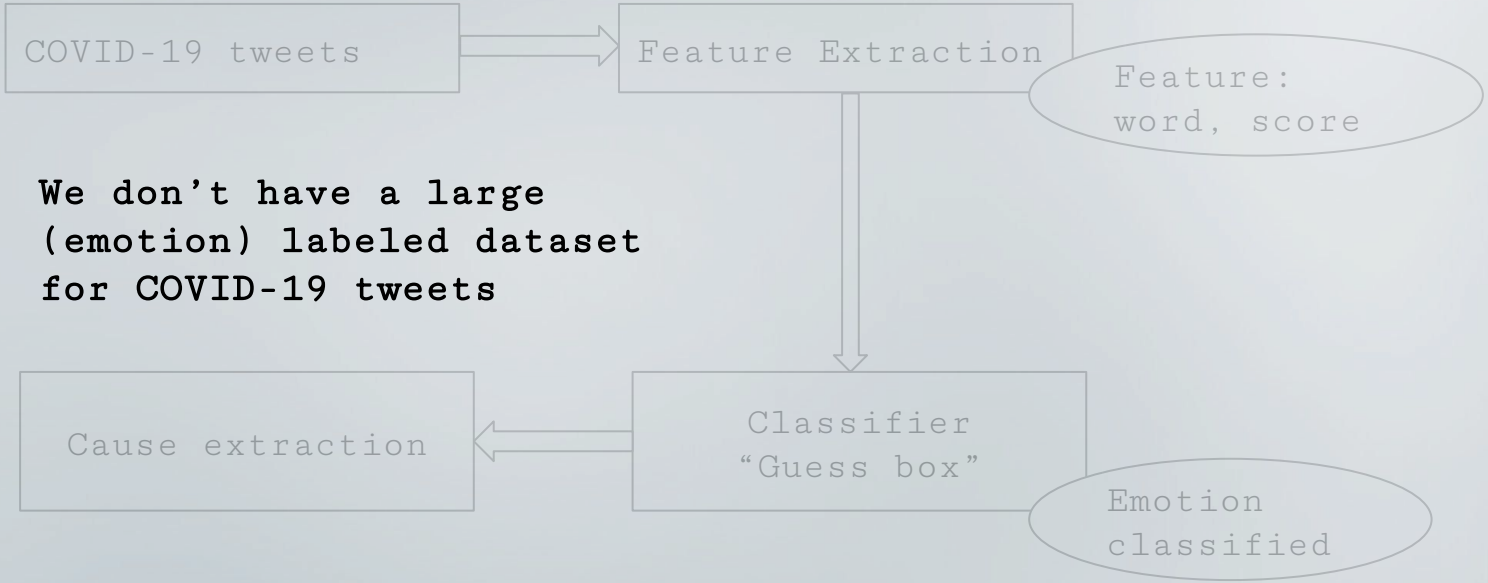
PIPELINE (initial)



PIPELINE (initial)



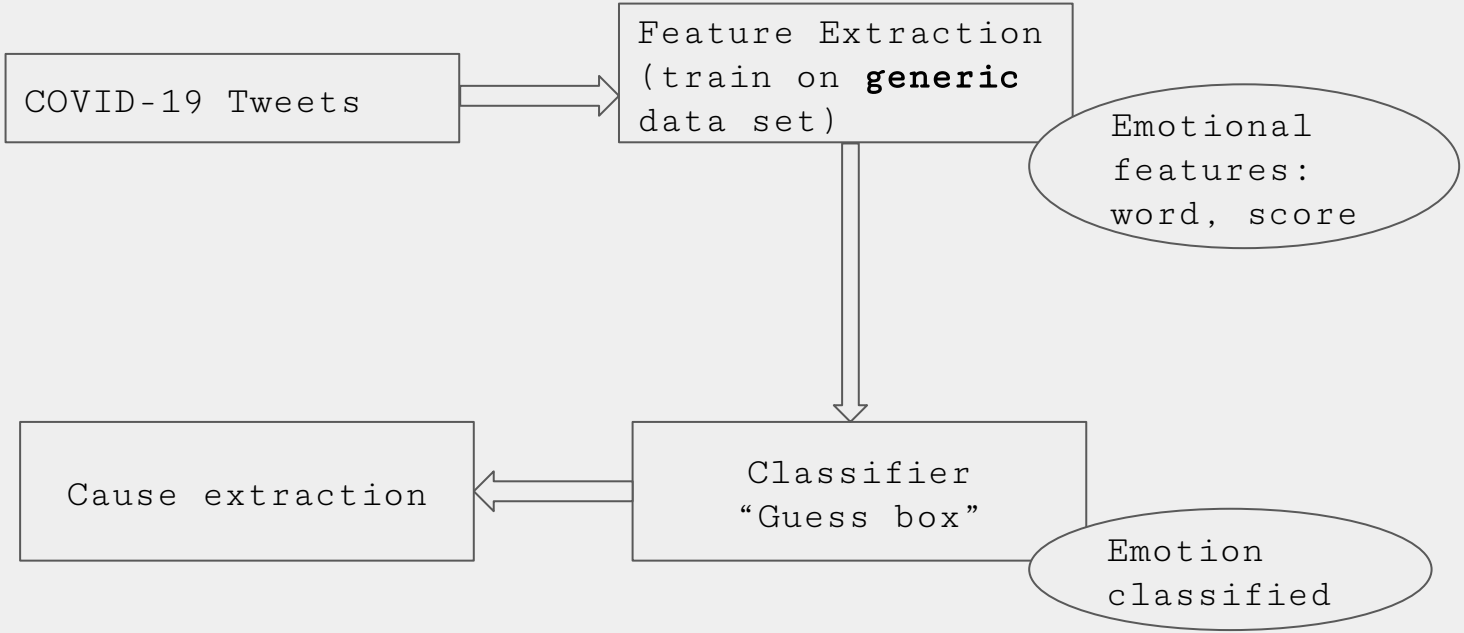
WHAT IS WRONG WITH THIS?



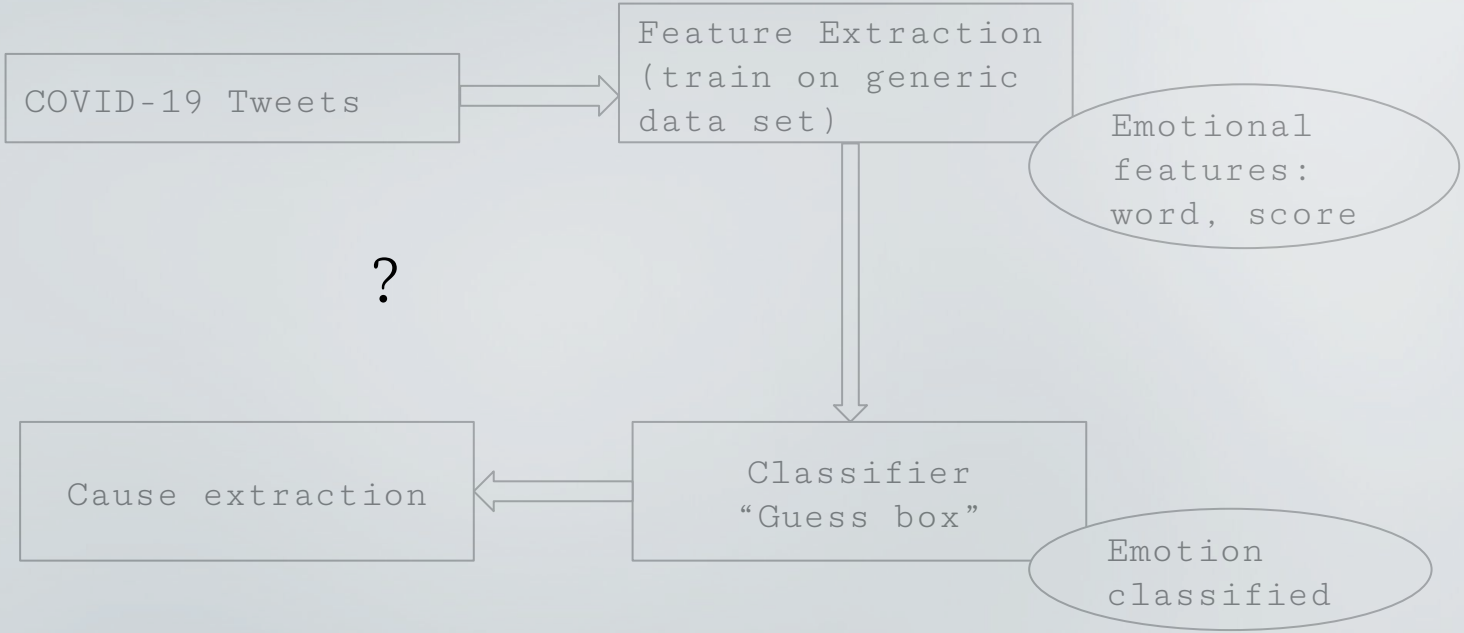
We don’t have a large (emotion) labeled dataset for COVID-19 tweets

SOLUTION: TRANSFER LEARNING

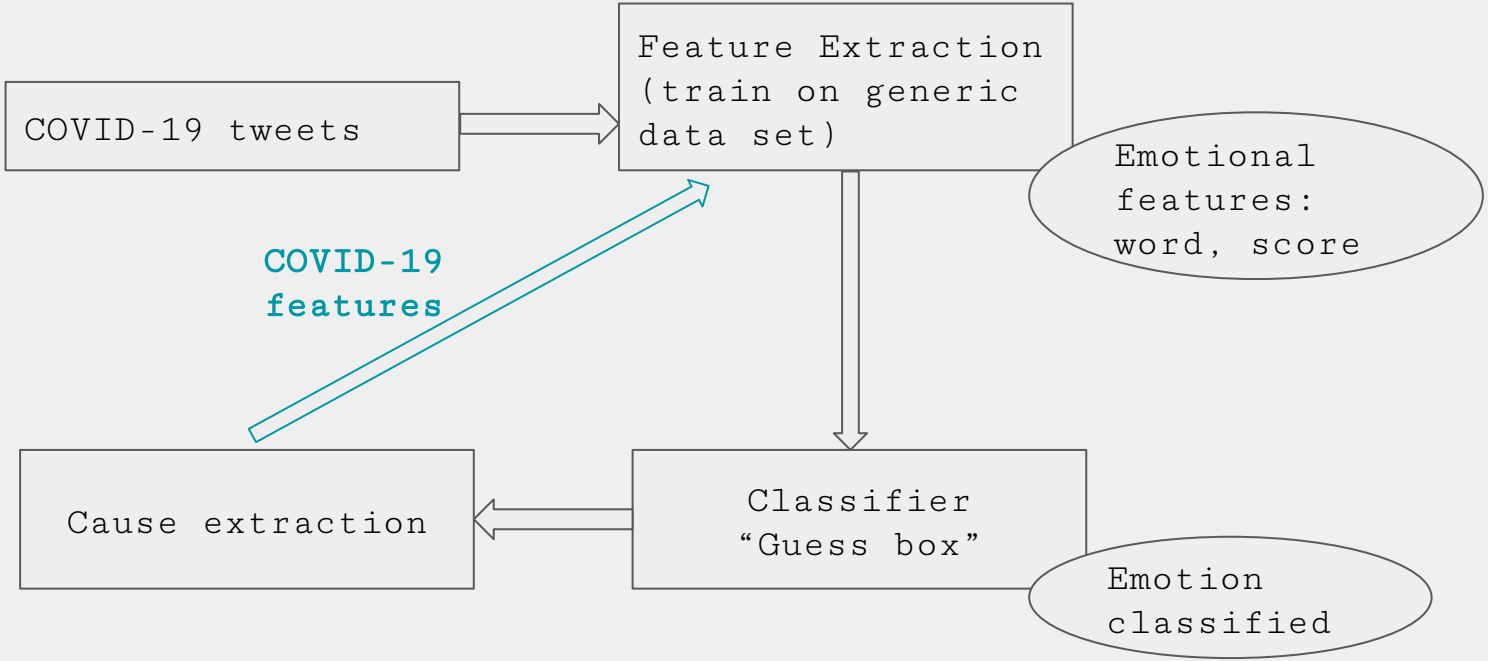
REVISED PIPELINE



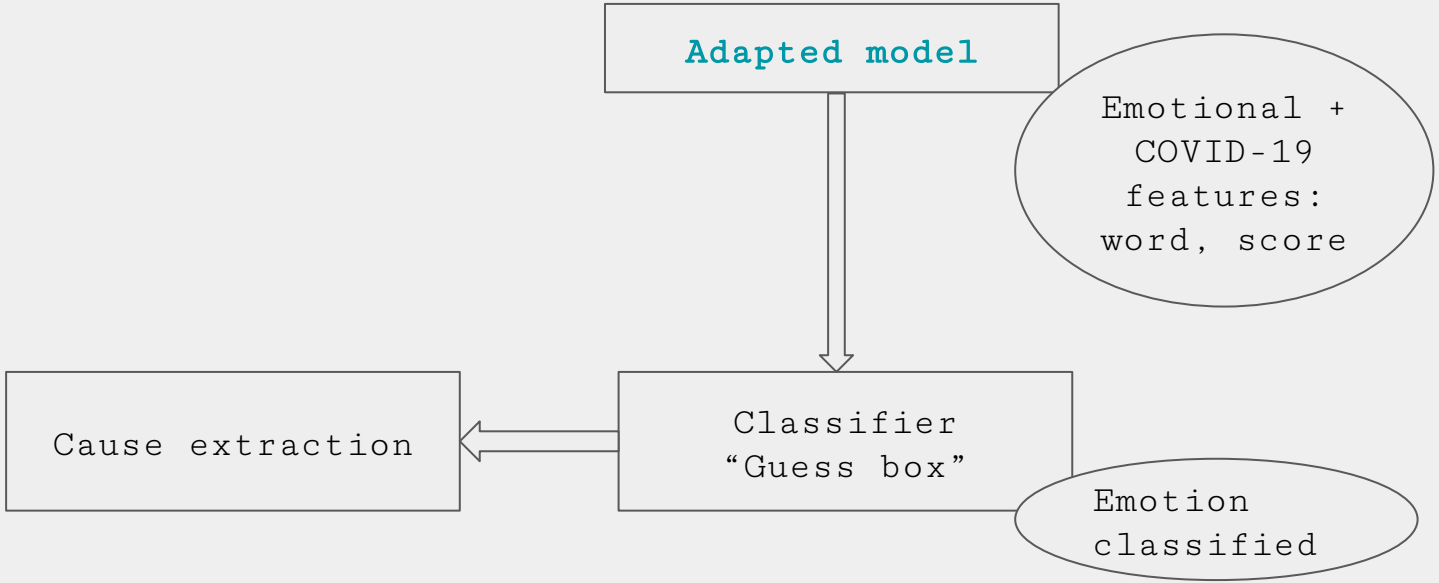
WHICH COVID-19 FEATURES TO ADD?



PIPELINE (updated)



PIPELINE (updated)



2. IMPLEMENTATION

DATASET

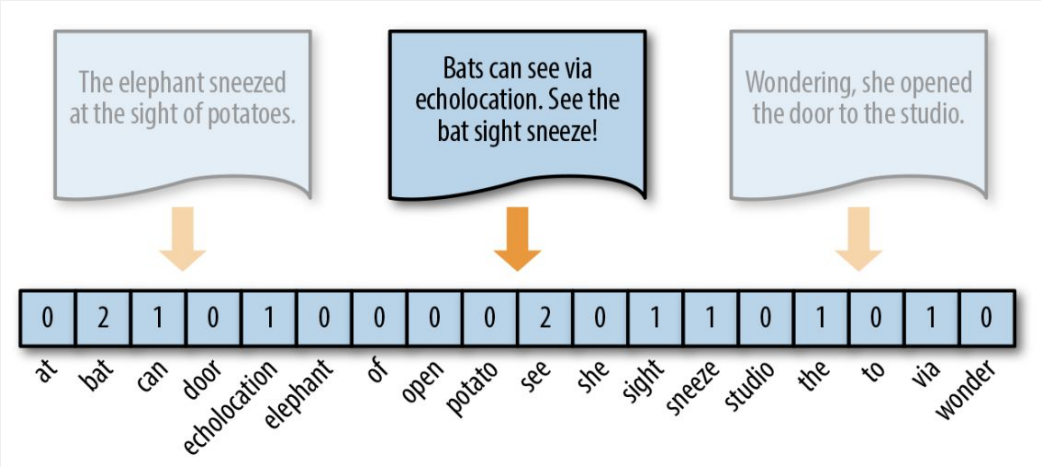
- **Target dataset: COVID-19 tweets** dataset on **Kaggle**

- **Generic dataset: Saif Mohammad** pre-labeled dataset
  http://saifmohammad.com/WebDocs/EmoInt%20Train%20Data/
  Use (75/25) train/test splitting
  Anger: 857, Fear: 1147, Joy: 823, Sadness: 786

- **COVID-19 mini-dataset: manually-labeled** tweets crawled from **Tweepy** by **keyword**
  Only for testing
  Anger: 25, Fear: 25, Joy: 25, Sadness: 25

IMPLEMENTATION: FEATURES EXTRACTION

- **Input: Labeled** tweets dataset
- **Output:** Emotional **features**
- **Tweets representation: Bag-of-words** (the occurrence of words within a tweet)



IMPLEMENTATION: FEATURES EXTRACTION

- **Input: Labeled** tweets dataset
- **Output:** Emotional **features**
- **Tweets representation: Bag-of-words**
- **Vectorizer**: **TF-IDF** Vectorizer

$$TF = Term\ Frequency$$
$$IDF = Inverse\ Document\ Frequency$$
$$TF\text{-}IDF = TF \times IDF$$

- **Features**: **Emotional words**

- **Scoring:** The score of **feature (word) w** for **class c** is based on the **log likelihood** that the **word w** appears in **class c** tweet

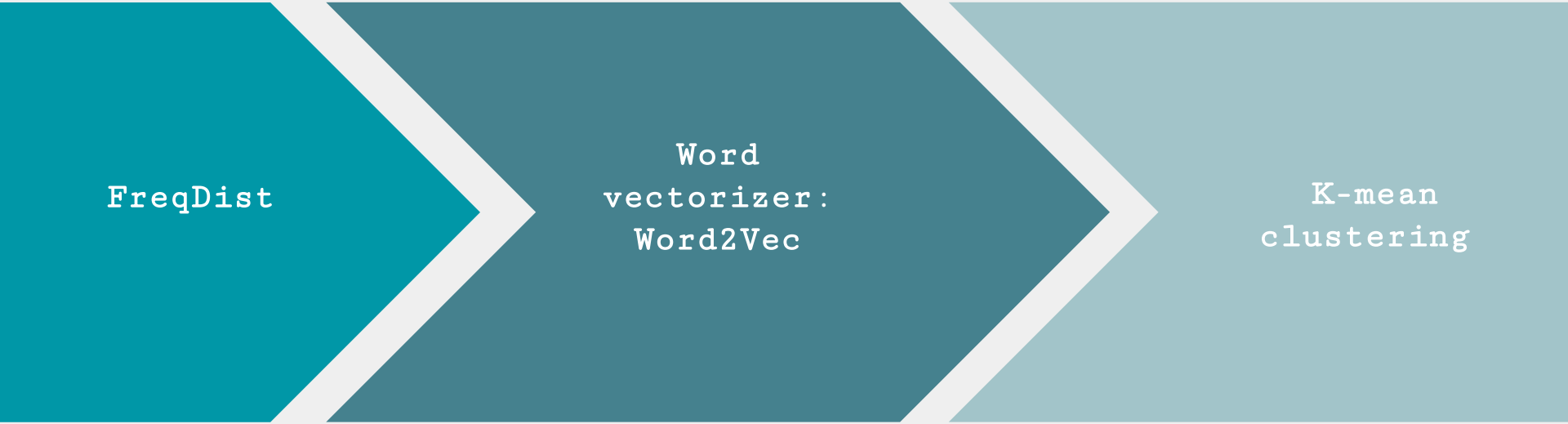
IMPLEMENTATION: CLASSIFIER

- **Input: Unlabeled** tweets
- **Output:** Emotional **class**
- **Tweets representation: Bag-of-words**
- **Vectorizer**: **TF-IDF** Vectorizer

$$TF = Term\ Frequency$$
$$IDF = Inverse\ Document\ Frequency$$
$$TF\text{-}IDF = TF \times IDF$$

- **Scoring: The score of a tweet is the sum of score of all words in that tweet**
- **Classification: The class with highest score**

IMPLEMENTATION: CAUSE EXTRACTION



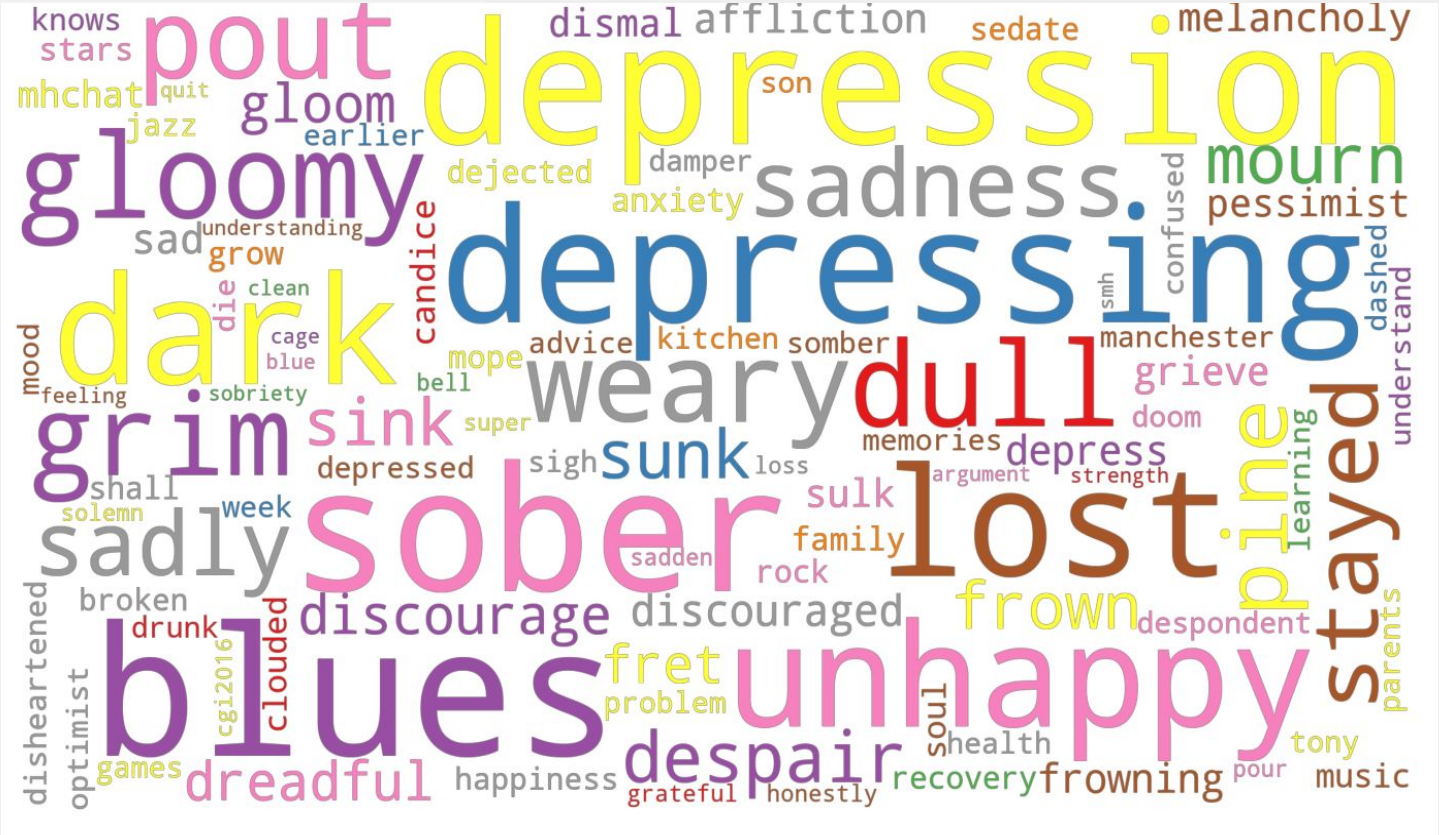
IMPLEMENTATION: FEEDBACK

- **Feature construction**: causes exclusively for one class, cause appear in 3 but 1 class, ...
- **Fine-tune**: Give **additional score** for the **occurence** of these **features** in the tweet
- The score given is **hyperparameter** (a.k.a we need to **fine-tune** it)

3. RESULT

RESULT: FEATURE EXTRACTION

Word Cloud for
SADNESS

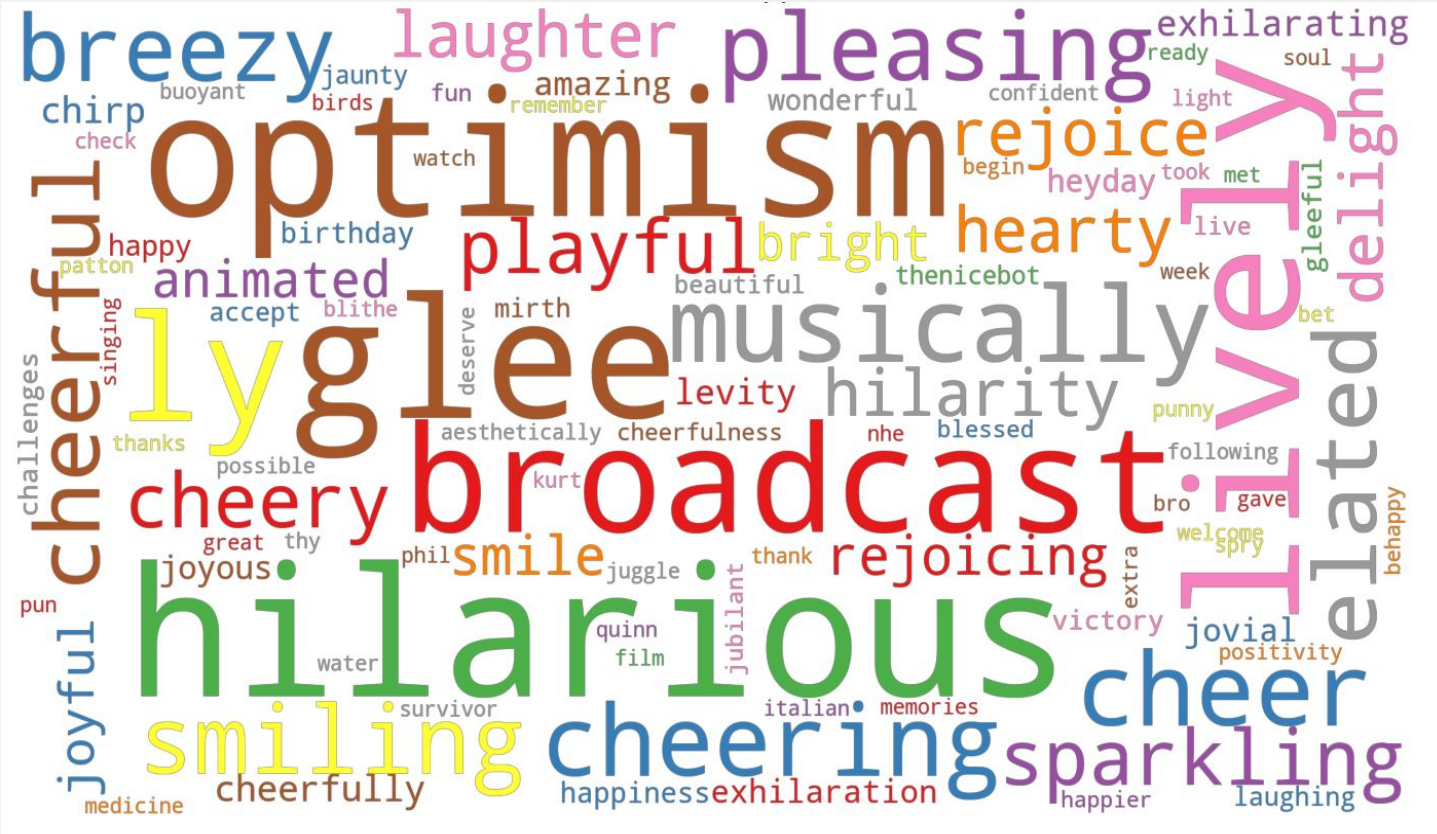


RESULT: FEATURE EXTRACTION

Word Cloud for **FEAR**

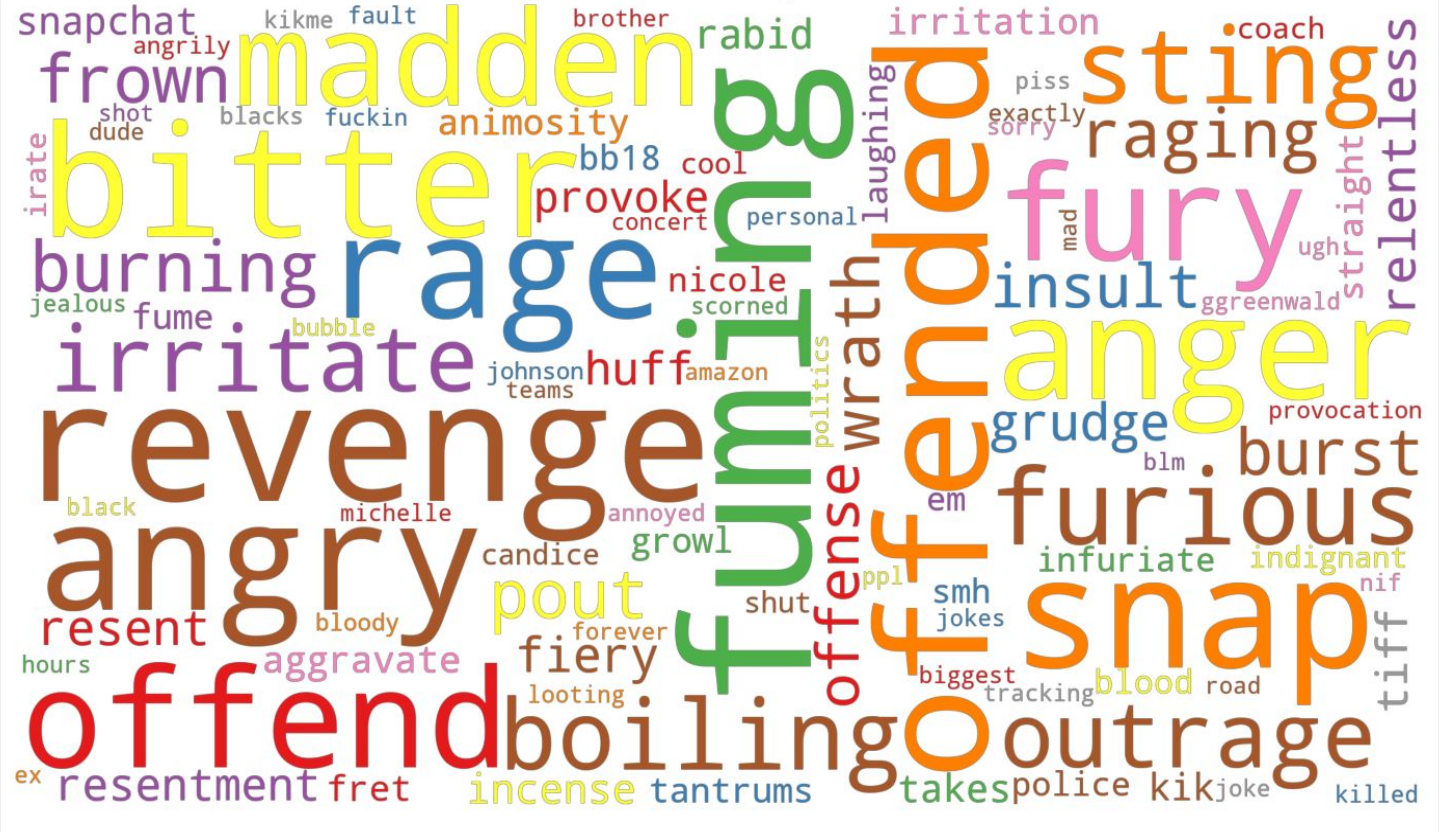
RESULT: FEATURE EXTRACTION

Word Cloud for JOY



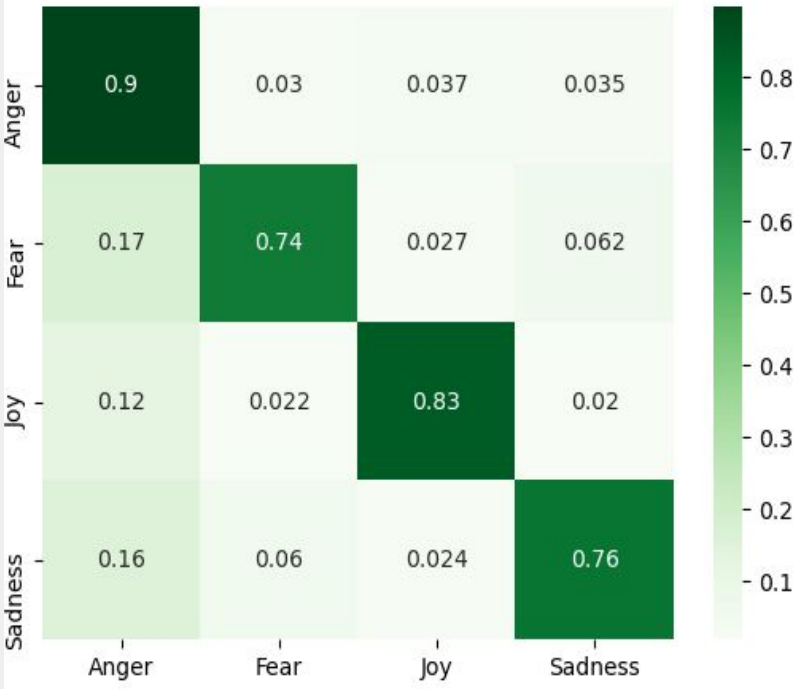
RESULT: FEATURE EXTRACTION

Word Cloud for
ANGER



RESULT: GENERIC MODEL (On general Tweets dataset)

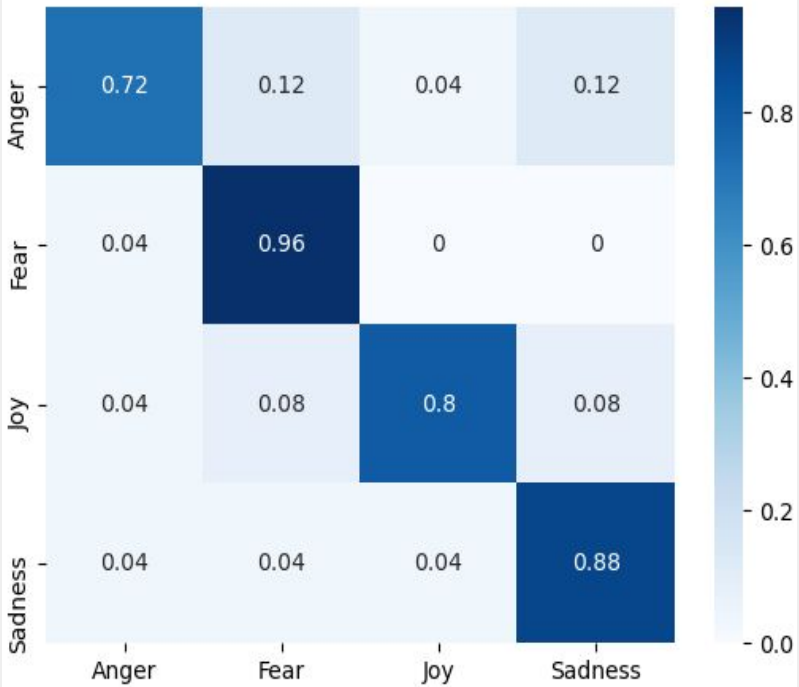
Accuracy: 80%



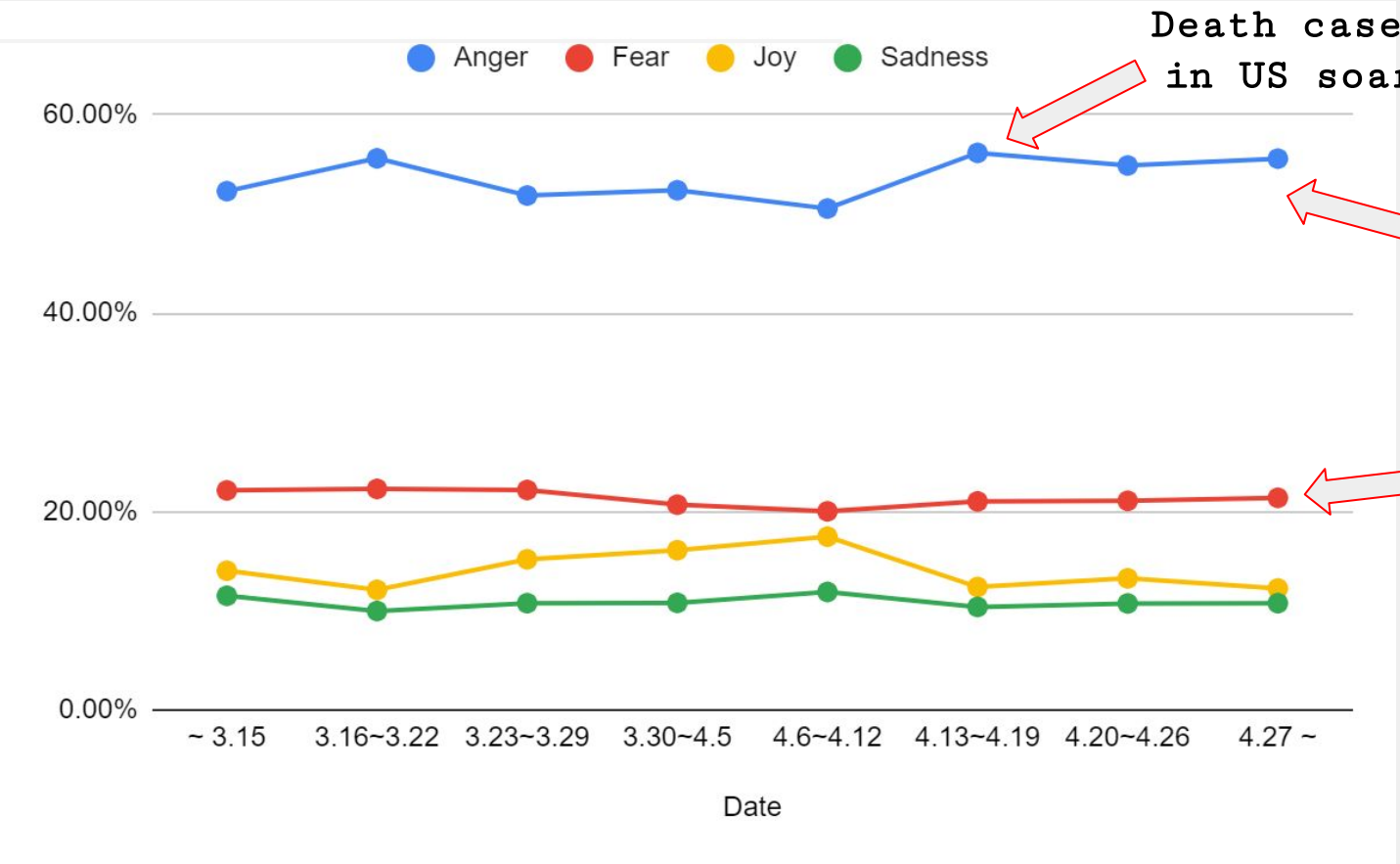
RESULT: ADAPTED MODEL (On Covid19 Mini-dataset)

Before Feedback: 83%

After Feedback: 84%



RESULT: Proportion of Emotions in Tweets



Death cases in US soar

ANGER is dominant

FEAR comes next

RESULT: example top anger tweets(4.13~4.19)

```
(0, 2.0946398357923597, 'This makes me so angry I can't
even express how angry I am #COVID19
#TrumpOwnsEveryDeath #RacismIsAVirus
https://t.co/fQTHFHdKhu')
```

AJ+ @ajplus · Apr 14

A Black doctor in Miami who has been testing the homeless for COVID-19 was detained outside his home while he was unloading his van.

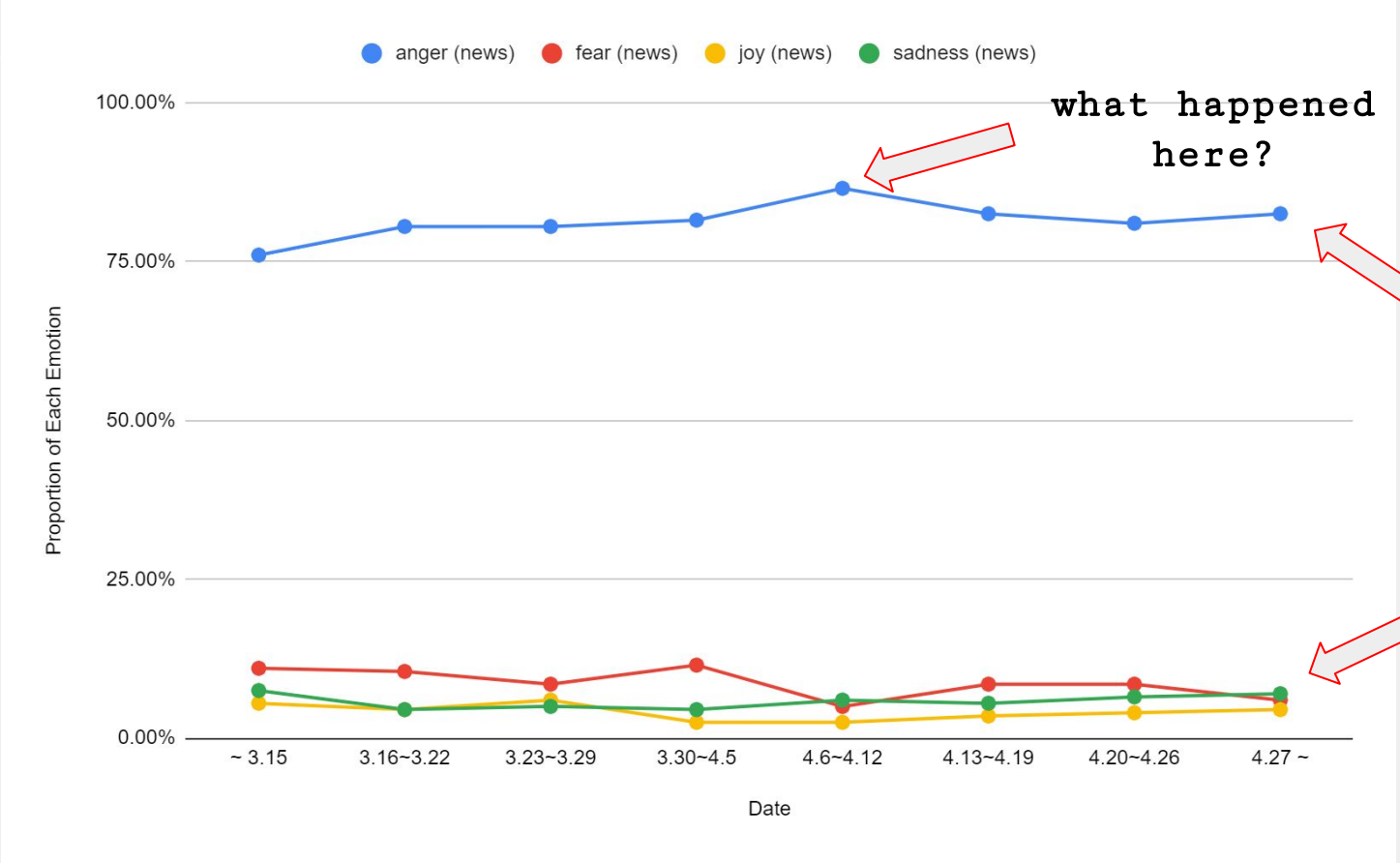
Police say they thought he was illegally offloading trash.

```
(0, 2.191388274890425, 'Unfucking believable! This is an
outrage in 2020! #Covid_19 https://t.co/tFFou5Hn1j')
```

Jeffrey Levin @jilevin · Apr 9

Two black men say they were kicked out of Walmart for wearing protective masks

RESULT: Proportion of Emotions in News



RESULT: example of top anger news(4.06~ 4.12)

S.Africa launches mobile tracking of those with coronavirus

2 MIN READ

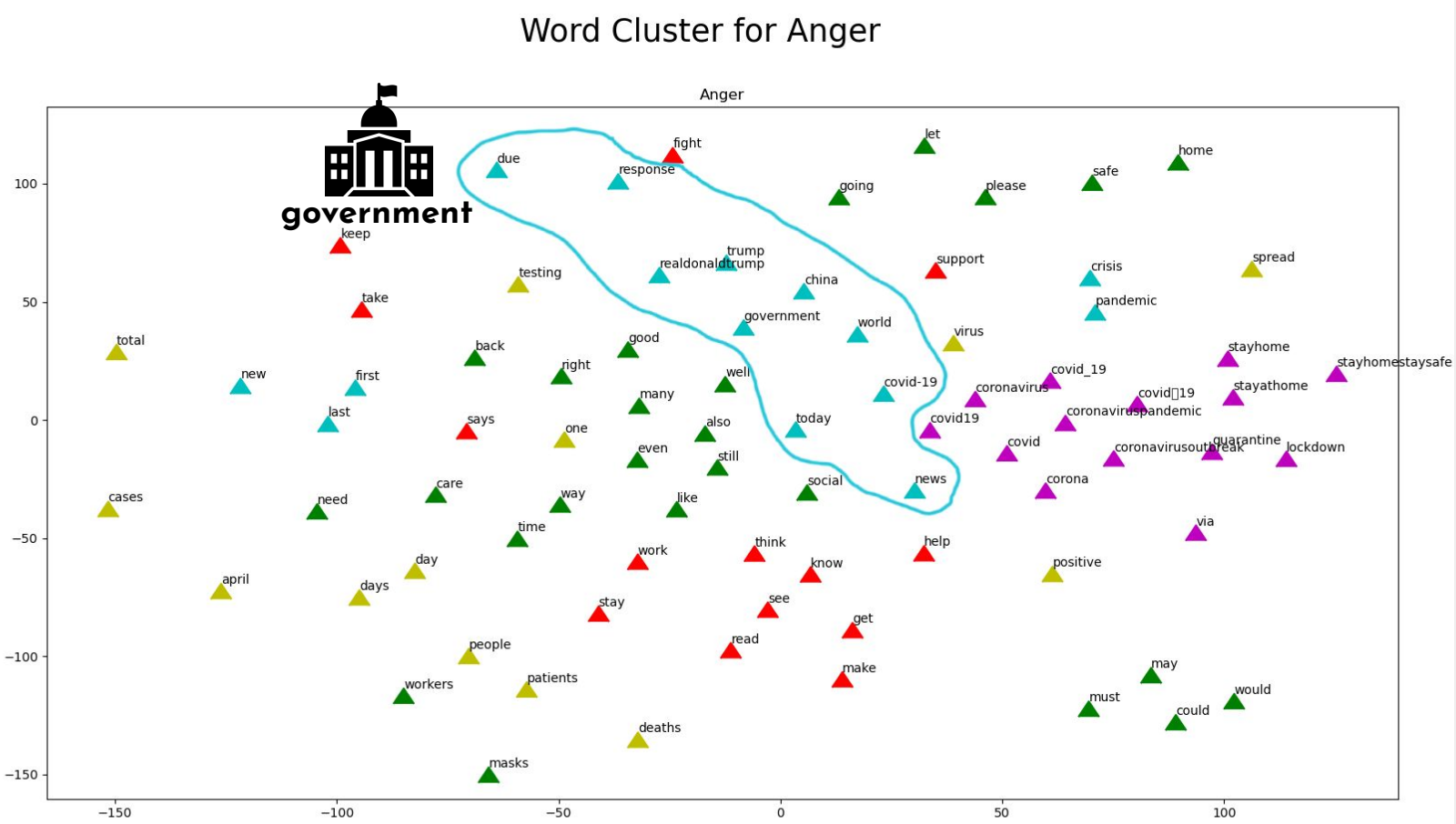
WHO'S LISTENING?

South Africans are worried the government will use coronavirus phone tracking to spy on them

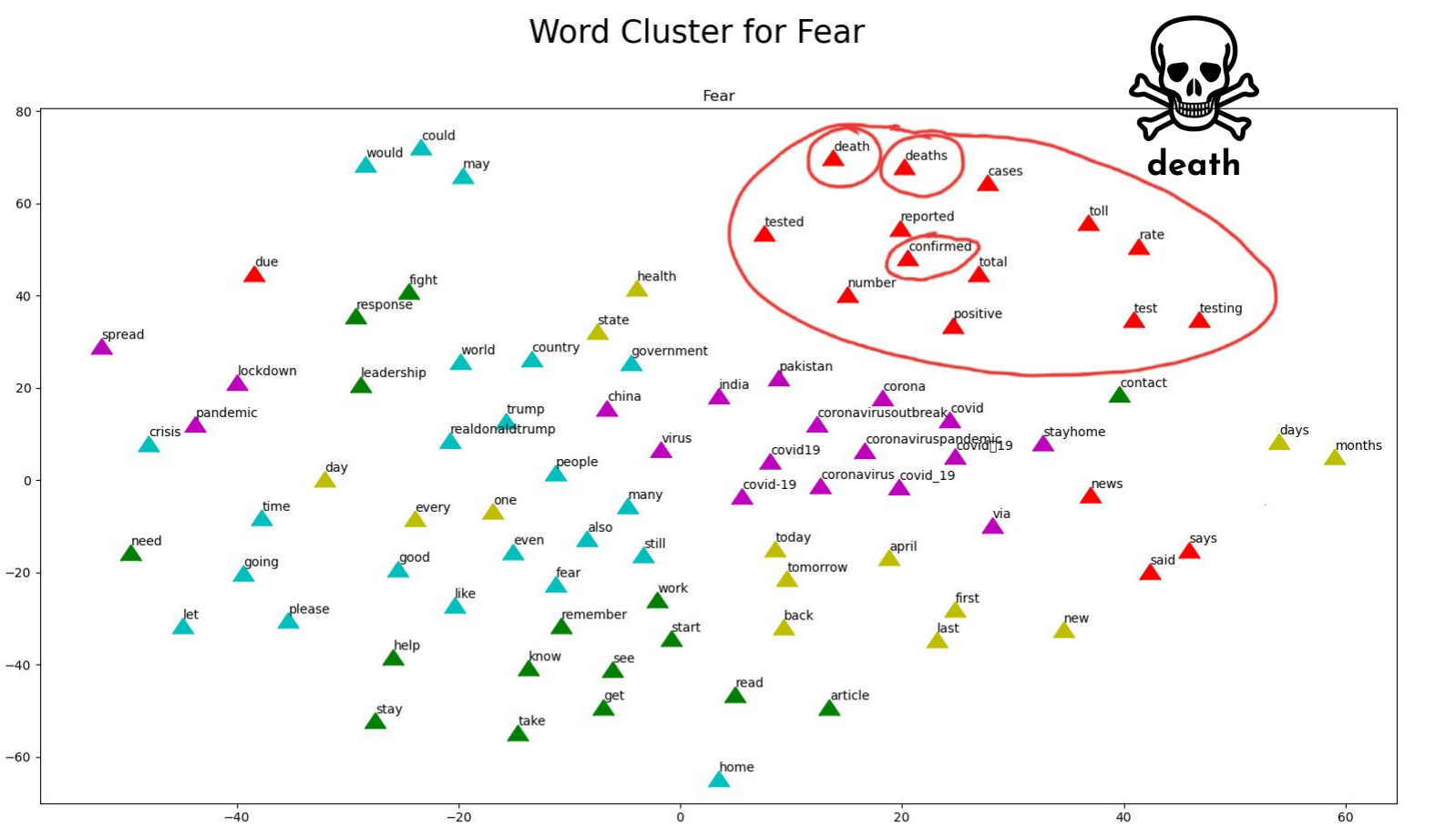
Johannesburg • April 8, 2020

By **Norma Young**

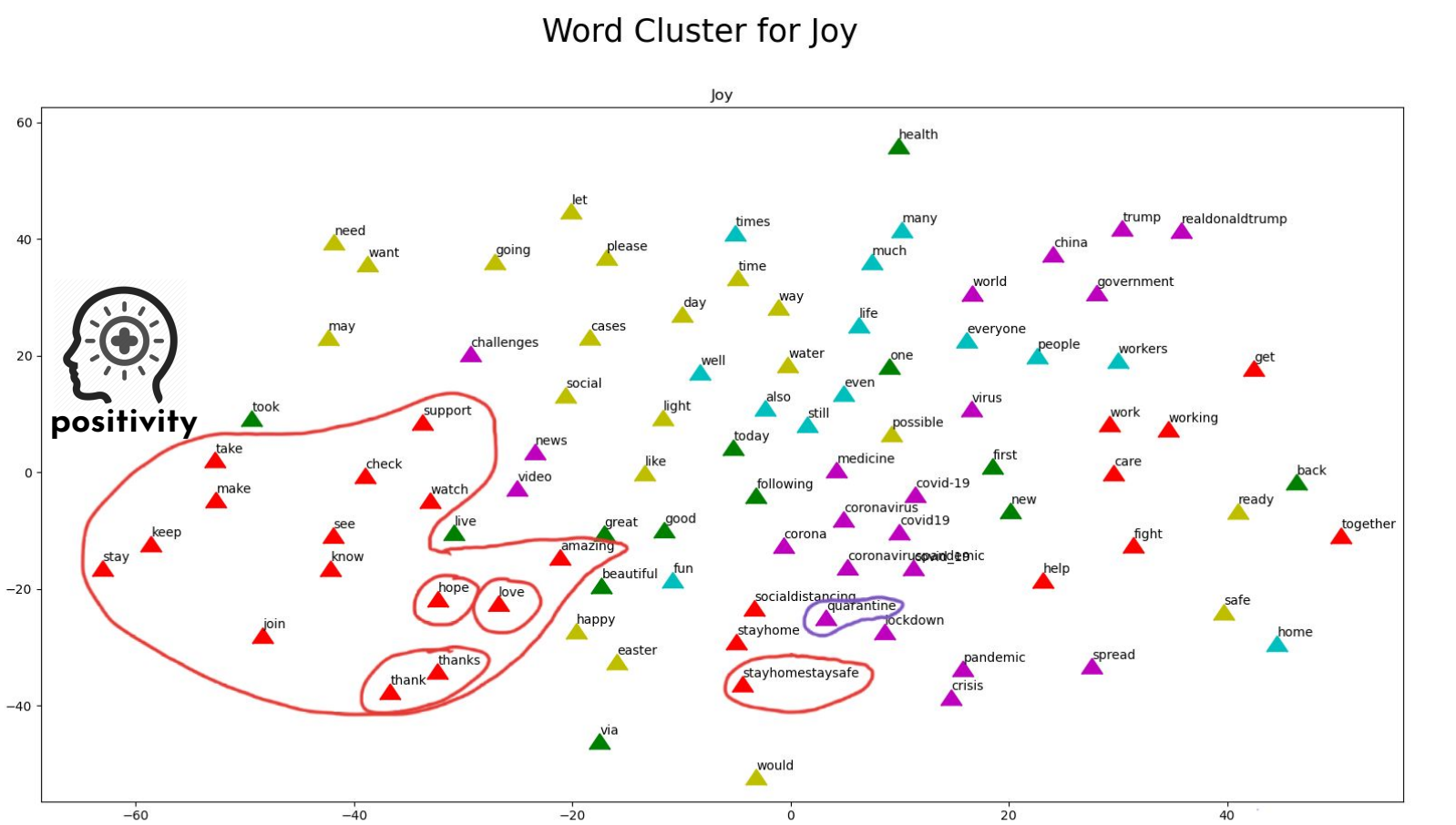
RESULT: CAUSE EXTRACTION (ANGER)



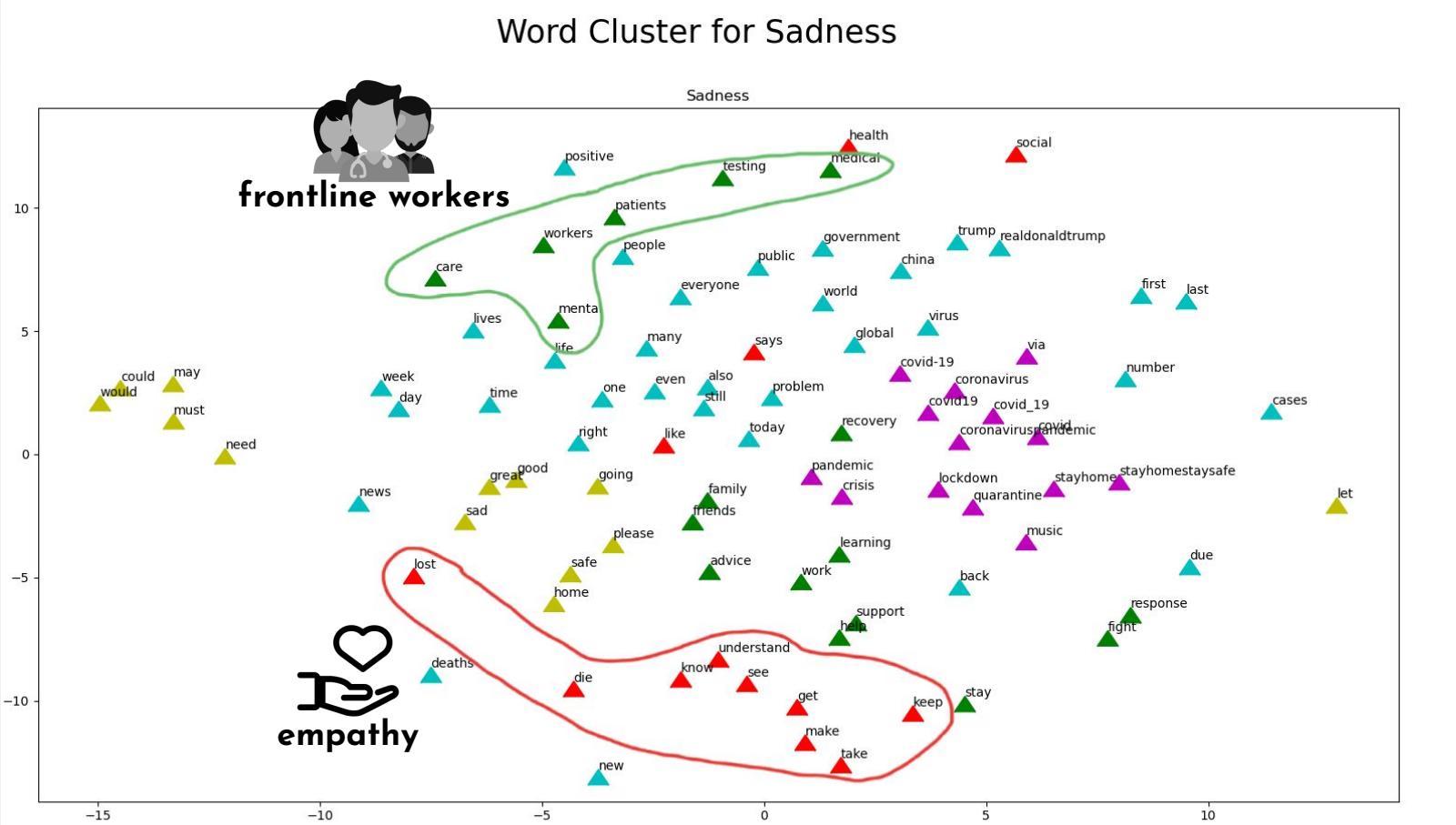
RESULT: CAUSE EXTRACTION (FEAR)



RESULT: CAUSE EXTRACTION (JOY)



RESULT: CAUSE EXTRACTION (SADNESS)



Dear Dr.,

Thank you for taking care of the patients. We appreciate you for taking the time out of your day to take care of everyone who is sick. Once again thank you.

4. CONCLUSION

PANicDEMIC Result Summary:

- Anger stays dominant.
- Causes of emotions: government response, quarantine etc.

Reflections About NLP:

- The features learning is important.
- Even with a simple model, the performance could be outstanding if we have good features.

Contribution / Application:

- Trend of people’s emotion during pandemic and their causes:
  - help in setting recovery strategy
  - Simple, interpretable model (Python, NLTK)

Further Improvement:

- N-gram cause extraction
- Analyze content of news (not only title)
- More locations and other languages