Declan Cassidy

Seungjun (Josh) Kim

Muhtasim Miraz

Depression during the COVID-19 Pandemic

Team “Pass or Fail”

The problem

Problem statement

Compared to the 2008-2009 Subprime Mortgage Crisis, are people feeling more depressed during the COVID-19 Pandemic?

Did U.S. government's responses quench people's anxiety / depression? If so, which policies / responses?

Data (1) - Oxford COVID-19 Government Response Dataset

Aims to track and compare government responses to the coronavirus outbreak worldwide rigorously and consistently.

Only subsetted the “U.S. Government Responses” because the focus of the analysis is towards U.S. users.

Variables are things like “school closure”, “workplace closure”, “income assistance” etc.

Codebook can be found here.

Data (2) - Tweet Data

- Tweets from 2009 (Kaggle) : Tweets from the 2008-2009 Subprime Mortgage Crisis

- Tweets with the following hashtags (early April, late April) :

#coronavirus, #coronavirusoutbreak, #coronavirusPandemic, #covid19, #covid_19, #epitwitter, #ihavecorona, #StayHomeStaySafe, #TestTraceIsolate

Tweets from the COVID-19 Pandemic

Data (3) - Reddit Comments

Referred to:

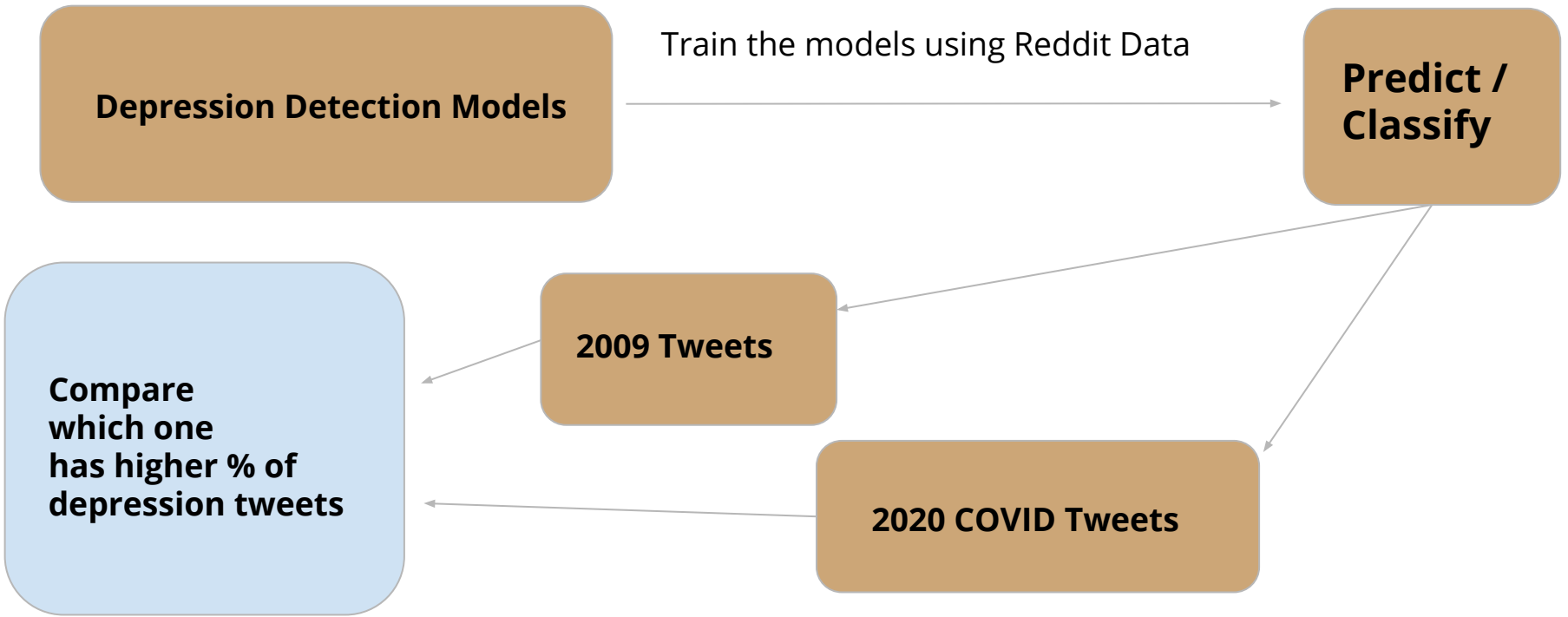
Wolohan “Detecting Linguistic Traces of Depression in Topic-Restricted Text: Attending to Self-Stigmatized Depression with NLP (2018)”
⇒ used Reddit Comments for Depression / Non-Depression Binary Classification Modelling.

subreddit “depression” (/r/depression) : y = depression

Subreddit “AskReddit” (/r/AskReddit) : y = non-depression

Dataset can be found here.

Methodology



Features

Features we extracted from text ⇒ 207 Features

word count
unique word count
stop word count
mean word length
median word length
character count
punctuation count
top 100 most frequent unigram ~ five-grams
top 100 unigram ~ five-grams based on tf-idf weighting

Results

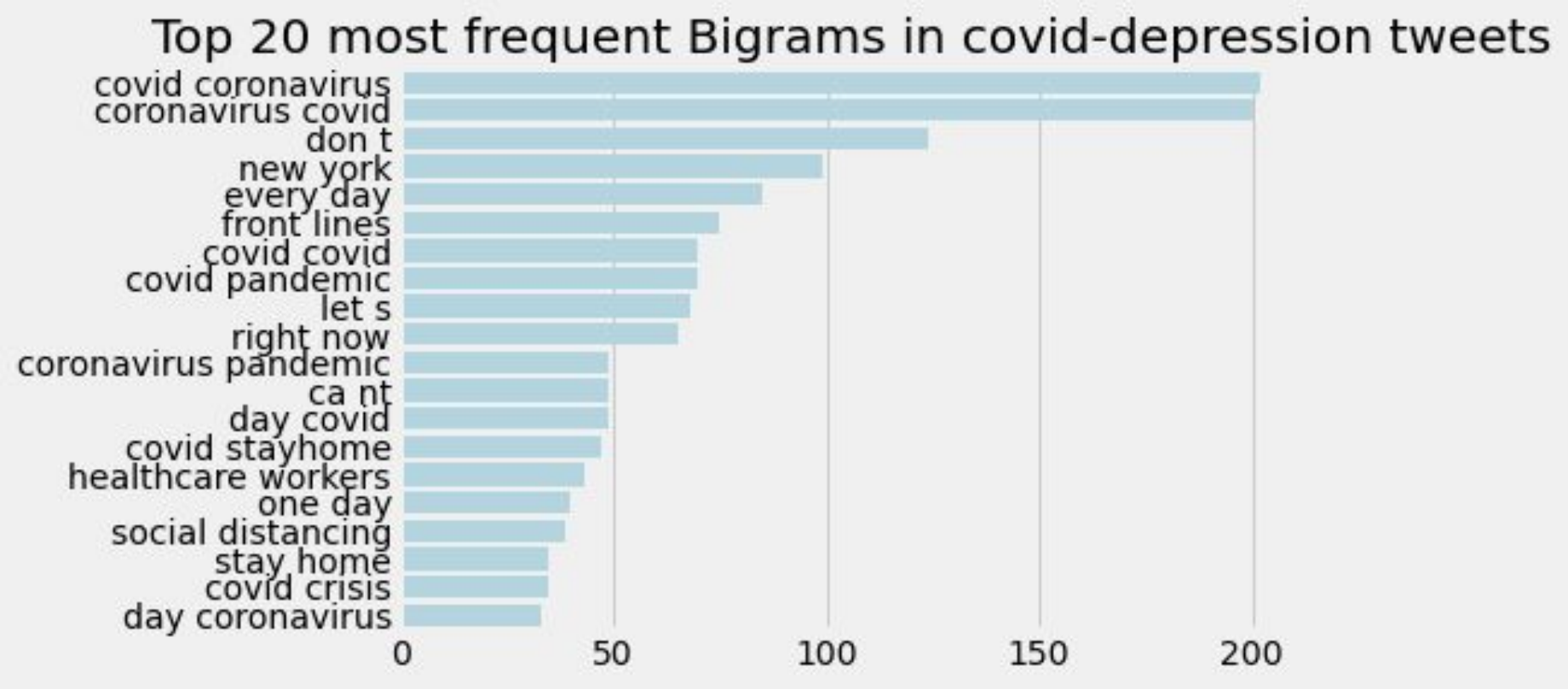
Subprime Mortgage Crisis was more depressing...?!

More people felt depressed during the 2008-2009 subprime mortgage crisis than the during the COVID-19 crisis.

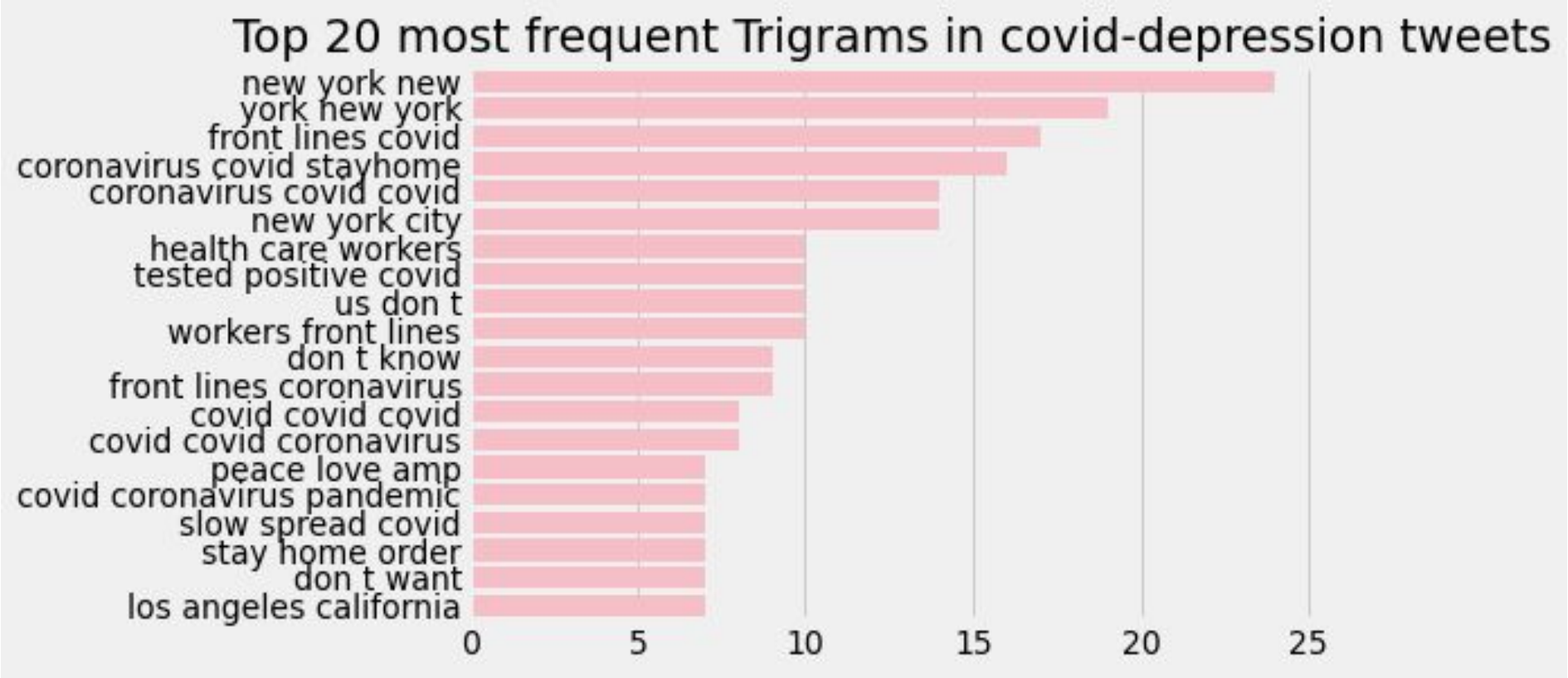
Depression Detection Model Performances using Reddit Comments Data & % of Predicted Depression

	Accuracy (CV = 5)	AUC Score (CV = 5)	F1 Score (CV = 5)	2009 Tweets	2020 Tweets
Gaussian Naïve Bayes	0.779	0.890	0.745	**9.16%**	**6.54%**
Logistic Regression	0.884	0.950	0.880	**26.98%**	**16.47%**
Random Forest (RF)	0.912	0.975	0.914	**17.92%**	**15.98%**
Light Gradient Boosting Machine (LGBM)	0.931	0.980	0.931	**20.16%**	**17.92%**
Majority Voting				**19.46%**	**15.42%**

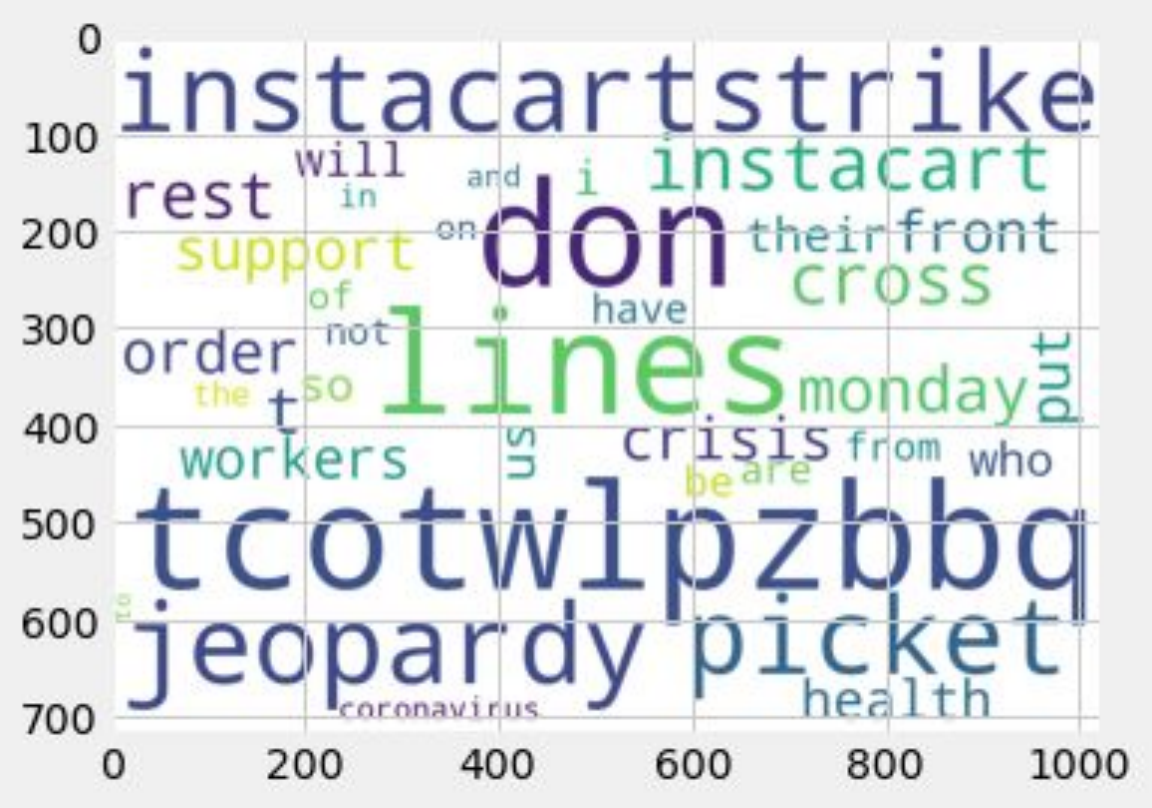
What words / phrases are associated with COVID-Depression?



What words / phrases are associated with COVID-Depression?



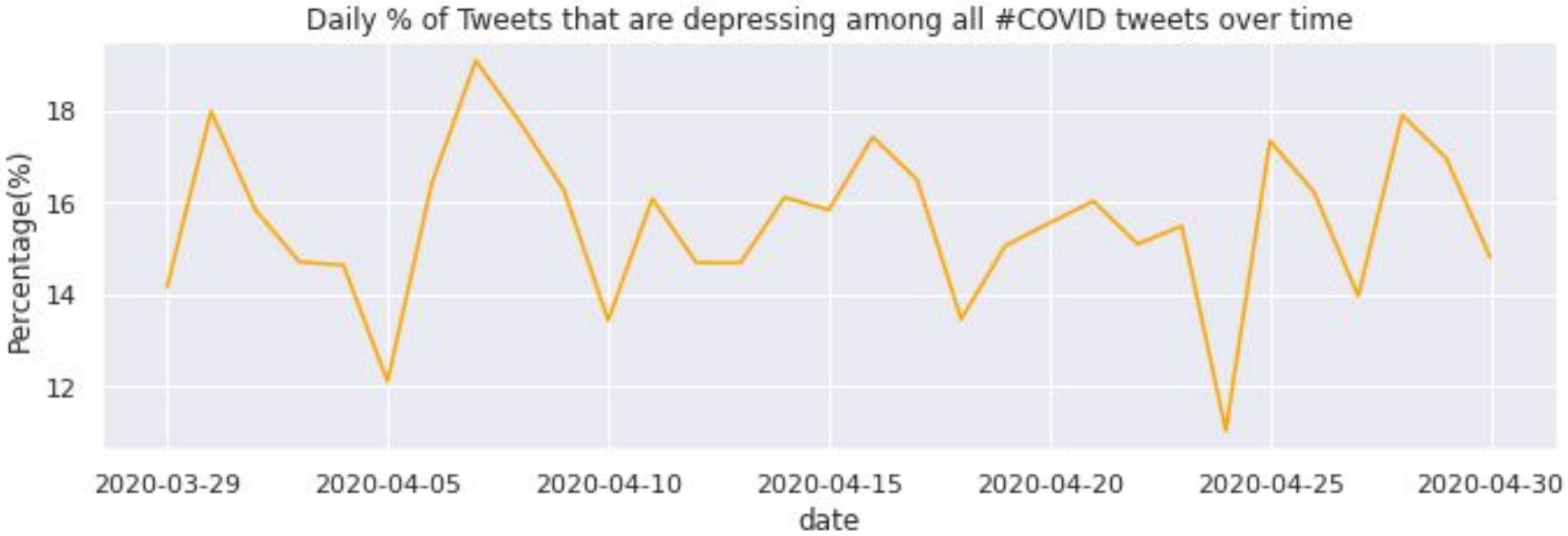
What words / phrases are associated with COVID-Depression?



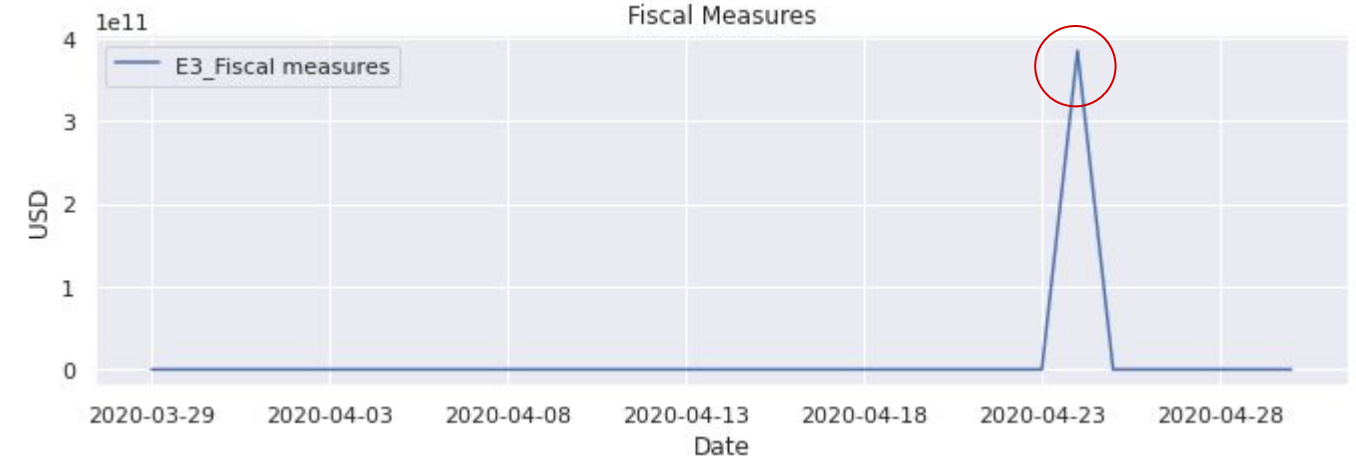
Bonus

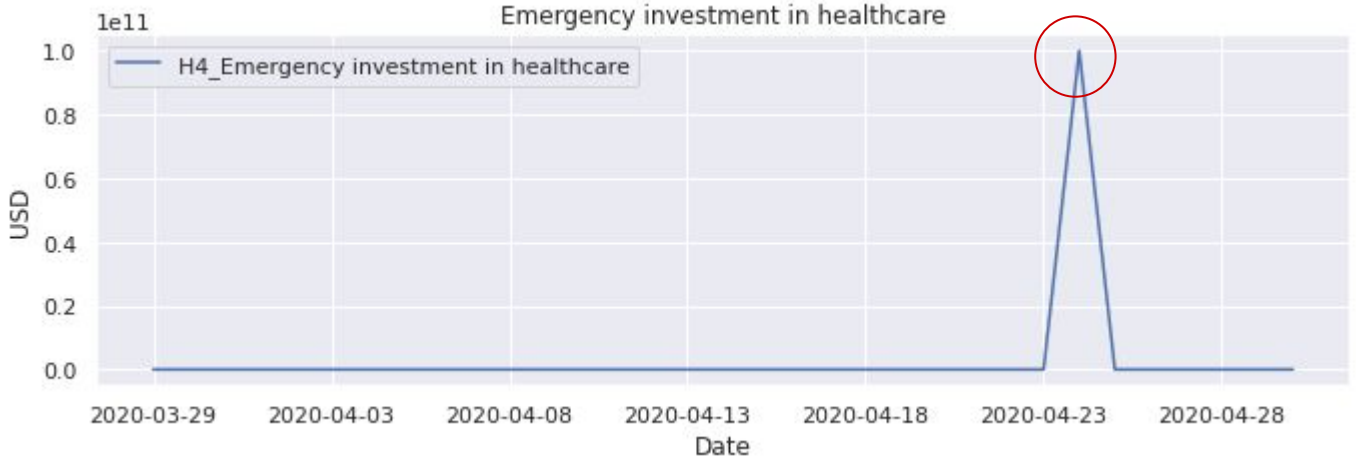
Were U.S. Government Responses Effective in Reducing Depression?

% of Depression Tweets among all COVID Tweets over time



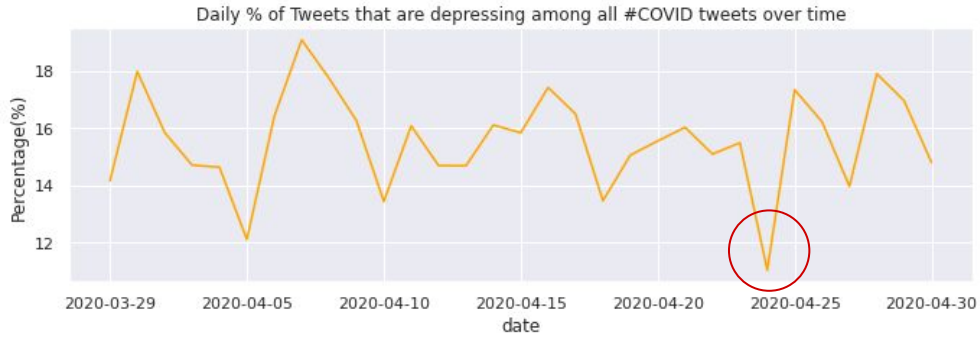
U.S. Government Fiscal Measures and Emergency Fund in HealthCare

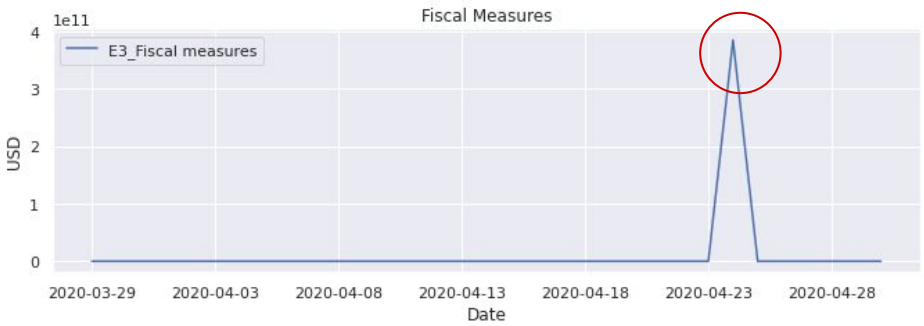


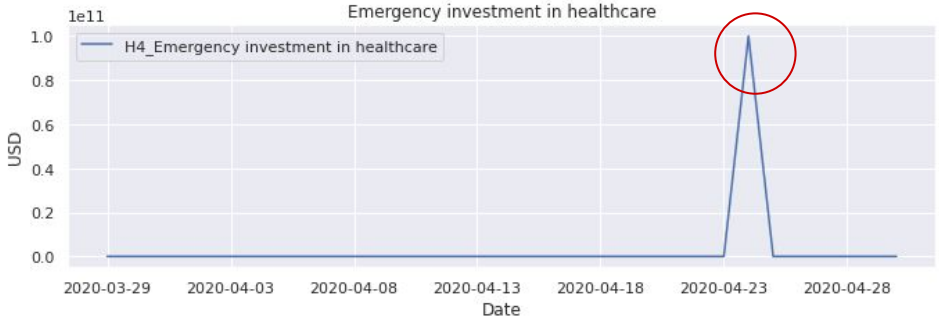


Drop in Depression Rate – Fiscal Measure, Health Emergency Fund

Association?







Limitations and Future Work

- We only random sampled part of the entire tweets due to computational reasons, so results may be biased. In the future, we can use more or all the data for robustness checks.
- If you want to gauge the impact of COVID on depression rates, it will be more accurate to compare between the tweets right before COVID started and the tweets after COVID spread to the entire country.
- We would like to try other feature representations (e.g. Word2Vec, Doc2Vec, POS Tagging) or use pre-trained models including the BERT model to see if the results we found still hold.
- The association between decrease in depression rates and fiscal measures may be due to other confounding factors. We would like to use more robust methods to see if there is a causality, not just an association.

Declan Cassidy

Seungjun (Josh) Kim

Muhtasim Miraz

Thank you!

Team “Pass or Fail”