Depression Levels of People during 2008 Subprime Mortgage Crisis v.s. 2020 COVID-19 Pandemic

Declan Cassidy, Seungjun (Josh) Kim, Muhtasim Miraz

Compared to the 2008-2009 Subprime Mortgage Crisis, did people feel more depressed? Did U.S. government's responses quench people's anxiety / depression? If so, which policies / responses?

Data

• Oxford COVID-19 Government Response Dataset: Aims to track and compare government responses to the coronavirus outbreak worldwide rigorously and consistently. Codebook can be found here.

• Tweet (U.S. Users) and Reddit Data

- Tweets from 2009 (Kaggle)
- Covid early and late April tweets from Kaggle (#coronavirus, #coronavirusoutbreak, #coronavirusPandemic, #covid19, #covid_19, #epitwitter, #ihavecorona, #StayHomeStaySafe, #TestTraceIsolate; early April, late April)
- Reddit Comments from subreddit "depression" and "AskReddit" sections

Referred to Wolohan's paper "Detecting Linguistic Traces of Depression in Topic-Restricted Text: Attending to Self-Stigmatized Depression with NLP (2018)" which used /r/depression comments as y = depression and /r/AskReddit comments as y = non-depression for ML.

Results and Findings

• Features we extracted (total of 207): 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

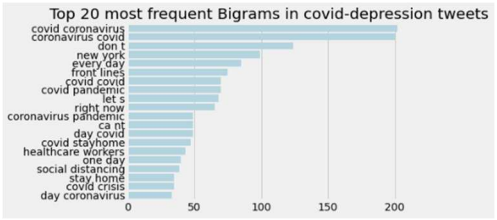
•

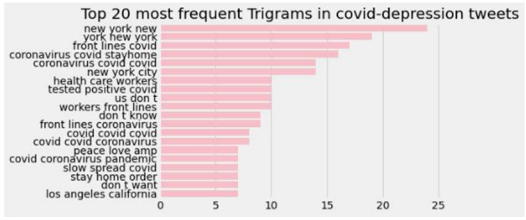
Using the features above, I trained the model and predicted / classified the 2009 and 2020 tweets into depression and non-depression.

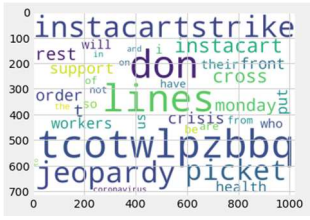
	Accuracy (CV[1] = 5)	AUC Score (CV = 5)	F1 Score (CV = 5)	**2009 Tweets**	**2020 Tweets**
Gaussian Naïve Bayes	0.779	0.890	0.745	**9.16%**	**6.54%**
Logistic Regression	0.884	0.950	0.880	**26.98%**	**16.47%**
Random Forest (RF)[2]	0.912	0.975	0.914	**17.92%**	**15.98%**
Light Gradient Boosting Machine (LGBM)[3]	0.931	0.980	0.931	**20.16%**	**17.92%**
Majority Voting[4]				**19.46%**	**15.42%**

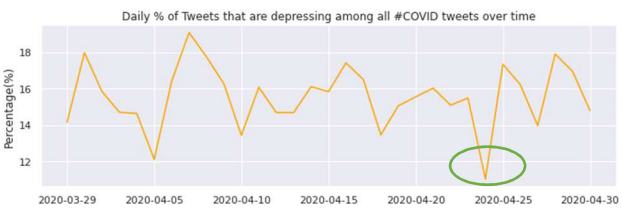
Across all models, the percentage of tweets that indicate depression is lower in the 2020 tweets than that in 2009 tweets. This means that more people felt depressed during the 2008-2009 subprime mortgage crisis than the during the COVID-19 crisis.

• Top 20 bigrams and trigrams for the COVID-19 tweets that were classified as "depressed" tell us that depression is associated with concerns towards front line workers, inability / lethargic state (e.g. can't, don't), dwindling hope for the status quo and the future (e.g. right now, every day) and high death rate (e.g. New York City: City with the highest death rate).

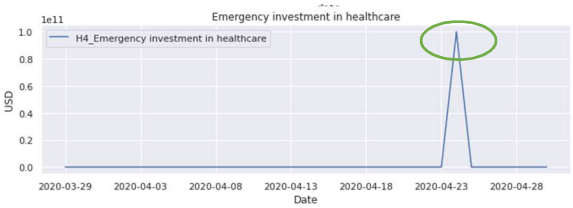


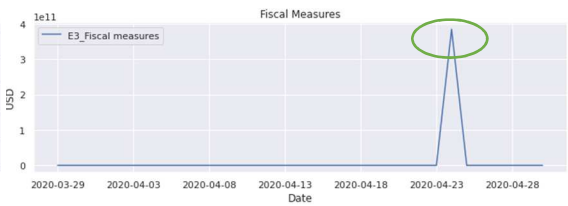






• Only fiscal measures (e.g. economic stimulus spending) and emergency health care spending (e.g. Announced short term spending on healthcare system, hospitals, masks) seemed to be associated with the decrease in depression. Other government responses (e.g. school shutdown) did not seem to have much correlation with increase or decrease of depression levels.





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 data for robustness checks.

[1] Cross Validation.

[2] N_estimators = 1000.

[3] boosting_type='gbdt', n_estimators = 5000, learning_rate=0.03, max_depth=-1.

[4] If three or more classifiers voted for depression, then mark it as depression (1). If only one or non classifier voted for depression, then mark it non-depression (0). For ties, follow the best performing model which is the LGBM model.