Measuring the Spatio-Temporal Psychological Impact of Government Policies during Covid-19 pandemic in different US states using Twitter Data

Vedant Khandelwal

I. Problem Statement

With over 5 million deaths and more than 262 million confirmed cases globally [1] by November 30th, 2021, COVID-19 has become a once in a century event. As there is little known about the novel coronavirus 2019, the uncertainty concerning the short and long term impact continues, threatening the well-being of individuals and societies at large. Further, considering the sudden stop of social and economic activities globally, the financial, addiction, and mental health implications have already started taking a toll on the US and its citizens. The introduction of specific rules to follow such as social distancing, stay-at-home, shelter-in-place orders, temporary closure of non-essential businesses led companies to lay off or furlough their employees, hiking the unemployment to around 16% from 3.6k% in just a month [2]. This sudden downturn in the economy affected people who were already isolated, increasing the propensity to use/abuse legal or illegal drugs and substances and develop anxiety and depression.

Building trust between the governed and the governed is critical to good governance. During the coronavirus epidemic, this claim has become a point of controversy. Compliance with any policy, especially during a pandemic, requires the public's support and trust towards the government. According to a survey in 2020, it is estimated that the public trust in federal government neat historic lows for more than a decade (less than 50%) [3]. Such a decline in trust can lead t0 lower compliance rates with new rules and regulations [4], [5].

II. Intended Users

There are two major intended users for the proposed work,

- **Policy Makers :** During the COVID-19, several policies taken by the state government were not timely, which led to a significant impact on the public's mental, social, and economic health. This work identifies such issues based on the public response towards the policy implementation [4], [6]. This insight on the response can help policymakers re-format their wheel of formulating policies and do a multi-disciplinary analysis with the help of psychologists, social scientists, statisticians [7], [8].
- **Public Health Specialists :** During the pandemic, the stay-at-home public gathering restriction orders led to a significant impact on people's mental health, as there was a significant increase in the mental health and domestic abuse hotline over time [9], [10]. This work helps identify and can be used to estimate the increase in mental health and addiction-related issues, hence helping the specialists prepare for the hike beforehand.

III. Dataset

Below is the list of datasets used for the proposed work:

- **Tweet-IDs:** These tweet IDs are used to hydrate tweets related to COVID-19. These tweet Ids are obtained from [], which contains an ongoing collection of tweet ids since 28th Jan 2020 [11].
- **Government Policies:** These are data related to several government policies implemented during the pandemic [12]–[15].
- **Diagnostic and Statistical Manual of Mental Disorders (DSM-5):** Also known as the bible for psychologists, this resource is used to obtain mental health and drug abuse-related entities [16].
- **Drug Abuse Ontology (DAO):** This ontology is used to obtain mental health and drug abuse-related entities and their usage in social media data and other platforms [17].
- **GeoNames Ontology:** This ontology is used to obtain granular location information about different US states, further used to tag different tweets with location data [18].
- **Dbpedia:** This is a general-purpose knowledge graph that enriches the lexicon of the mental health and drug abuse entities [19].
- **Subreddit:** These are posts and their top comments from Depression, Addiction and Anxiety subreddits [20].

IV. Proposed Approach :

The proposed approach for this work is shown in Figure 1. Different components of the approach are discussed below:

A. *Entities extracted from News Articles*

News articles provide retrospective information on an incident if reported on Twitter. However, the content is relevant without the need for further clarification. The content is free of grammatical errors, misspells, and misinformation as it undergoes curation from the human. The entities extracted from these are helpful to keep the set of lexicons updated with the use of different words in the context of mental and drug abuse.

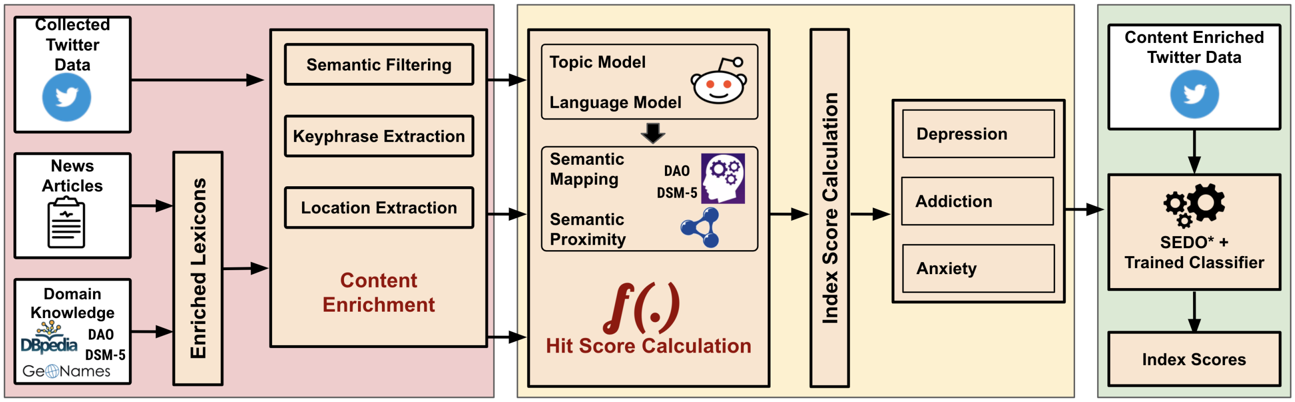

Fig. 1: Proposed Technical Approach

B. Enriched Lexicons

A set of lexicons is obtained from mental health and drug abuse entities extracted from DSM-5 and DAO. Further, these entities are enriched using the entities obtained from new articles and general-purpose knowledge graphs such as DBpedia.

C. Semantic Filtering

The tweet ids are obtained using the tweepy API based on the keywords related to the COVID-19. Obtaining a massive amount of tweets does bring in garbage tweets, which are not related to the COVID-19, even after supplying the required set of keywords. To remove the garbage tweets, we employ the method of semantic filtering and filter out the tweets based on the semantic distance between tweets and COVID-19 using cosine similarity.

D. Location Extraction:

During emergencies, the geographical location information of the events and that of affected users are vitally important. Identifying this geographic location is challenging, as available location fields such as user location and place names of tweets are unreliable. The location of users (e.g. state, city, and county) is rare in tweets and sometimes incorrect in terms of spatiotemporal information. This study utilized a simple but effective three-fold approach to location extraction from tweets showing potential mental illness signs. This would support local authorities and humanitarian organizations better to better situate their care strategies during this pandemic. We describe our approach as follows:

- We created a fine-grained location dictionary containing State, County, City, and Alias information using data.gov.us and Geonames ontology. The granularity of information in the dictionary allows location disambiguation while geo-parsing the tweet.
- We use ”geograpy” python API to extract location-specific information from the user’s content and metadata.
- To resolve the misspells, grammatical errors, and disambiguation, we employed a soft string matching approach between the tweet and location dictionary using the Levenshtein distance.

E. Keyphrase Extraction

A set of keyphrases are extracted from each tweet, which is further used in the analysis to calculate the hit scores. The key phrase is n-grams extracted from tweets, and these n-grams carry the context along with the important set of entities.

F. Topic and Language Modelling:

For this block, we perform topic and language modelling. To obtain the topics describing subreddits, we first use the skip-gram model to generate n-grams. Later we train LDA over subreddits and an LDA over bi-grams of subreddits. Relevant topics are identified by employing the topic coherence measure. We train a word2vec model over n-grams obtained from subreddits for the language model.

G. Hit Score Calculation

Two important part of Hit Score calculation is Semantic Proximity and Semantic Mapping.

- **Semantic proximity** refers to the alignment of tweets with MHDA lexicons. As shown here, palpitations and social anxiety can be aligned with the concept of anxiety in MHDA.
- **Semantic Mapping**, employs the usage of trained topic models and language models to match compound topics from subreddits to those from tweets.

This hit score calculation is used to obtain the Index score values for Depression, Addiction and Anxiety.

H. Training Classifier

We use the tweets and their index scores obtained for the mental health and drug abuse categories to train three different binary classifiers.

1) SEDO: Semantic encoding and decoding optimization (SEDO) [21], [22] is used to incorporate the domain knowledge in the form of enriched lexicons into the classification process. Using SEDO, a discriminative weight matrix is obtained, which is further used to module the word embeddings of tweets based on the proximity of the word to mental health and drub abuse lexicons. This process uses the Sylvester

Model	Depression			Addiction			Anxiety		
	Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Naive Bayes	84.85	82.68	85.17	82.74	80.46	84.15	82.53	81.87	80.64
Random Forest Classifier	91.98	91.81	91.87	90.02	90.36	91.98	90.76	92.78	91.46
Balanced Random Forest Classifier	92.32	92.43	92.01	91.53	91.78	**92.57**	**94.37**	**93.87**	**92.47**
Balanced Sub-sample Random Forest Classifier	**94.12**	**93.02**	**92.57**	**91.64**	**91.82**	91.97	93.46	93.85	91.75

TABLE I: Training Classifier Evaluation Metric Scores

equation, which has been used in computer vision within zero-shot learning. Once finished, it outputs a matrix, where every entity extracted from the tweet and which aligns with the mental health and drug abuse lexicon has a weight vector of length. Each dimension of the weight vector refers to Depression, Addiction and Anxiety, respectively.

2) Training Classifier: The matrix obtained from SEDO is further used to module the word embeddings given as input to the classifier and the index scores. We train Random Forest, Naive Bayes, Balanced Random Forest and Balanced Sub-sample Random Forest Classifiers.

V. Experimentation and Results

A. Training Classifiers

We have trained different sets of classifiers for depression, addiction and anxiety, the scores of which are given in Table I. We found Balanced Random Forest and Balanced Sub-sample Random Forest Classifier to perform the best out of all three models.

B. Social Quality Index (SQI)

A Social Quality Index (SQI) is calculated from the aggregation of mental health and addiction components. Raw SQI considers tweet concepts abstracted through three different mental health lenses in the MHDA-Kb: Depression, Anxiety, and Drug Abuse Disorders. Raw SQI aggregates the relevant features with respect to each of these lenses in each message and does not consider preceding state conditions. Change in SQI is also potentially informative, particularly for comparisons between states. We transformed raw state SQI into a relative state ranking to capture drifts between worsening and improving psychological conditions in social quality. SQI ranking is also used to examine the effect of external factors, such as school closure, business closure, unemployment, and lockdown (including the extension of lockdown).

C. Analysis Timeframe

In Figure 2, we can see the graph for 46 weeks of COVID deaths and cases. During the 46 weeks of the analysis timeframe from 14th March 2020 to 31st January 2021, I divided the timeframe into two major parts and performed analysis during those. I have chosen Part 01 and Part 02 for the analysis because of high cases and deaths during those intervals, hence going with the assumption that high cases and deaths will affect the mental health of the public and hence the social quality [23], [24].

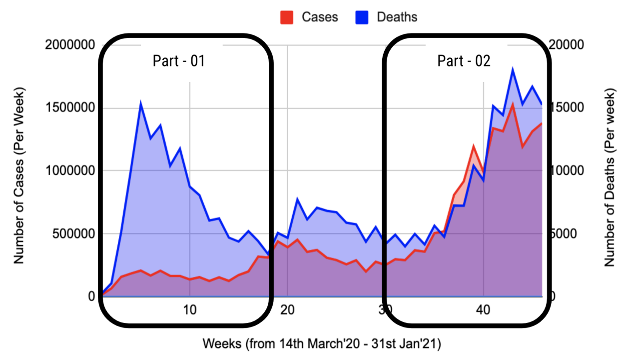

Fig. 2: US COVID cases and Deaths (14th March 2020 - 31st Jan 2021)

D. Part 01 Analysis (14th March 2020 - 20th June 2020

We have grouped several states to obtain clusters of states with a similar pattern of change in SQI for four weeks as shown in Figure 3. The graphs show the state's rankings, where the darker the shade of the state, the better the social quality. As we can see, the cluster of states consisting of New Hampshire, Ohio, Oregon, Washington, and Wyoming are worsening over the four-period timeline.

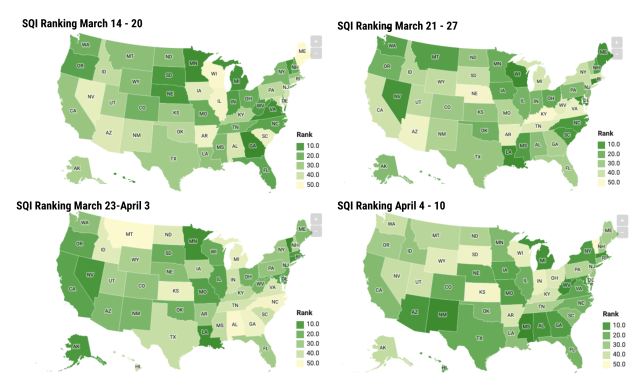

Fig. 3: Change of SQI in the first four weeks

We compared two clusters, an improving SQI cluster and a declining SQI cluster, and mapped various events and government policy implementation. A graph for improving SQI is shown in Figure 4, consisting of states: Georgia, Idaho, Indiana, Maryland, Maine, Missouri, North Dakota, South Dakota. Here, in this cluster, we can see that the financial policies such as federal funds for the unemployed, CARES Act fund release, extension of tax payment deadline, unemployment insurance funds, reopening of small businesses, paycheck protection programs and so on are having a positive

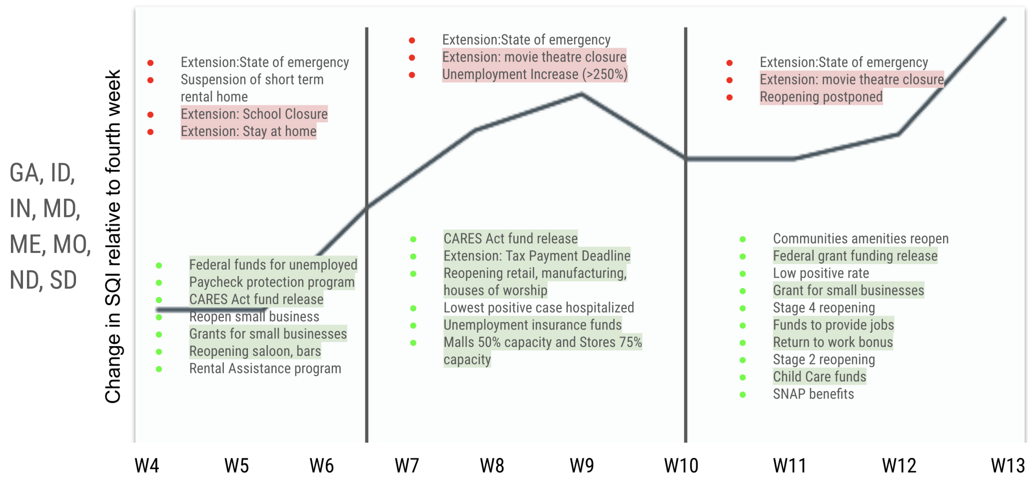

Fig. 4: Improving SQI Cluster

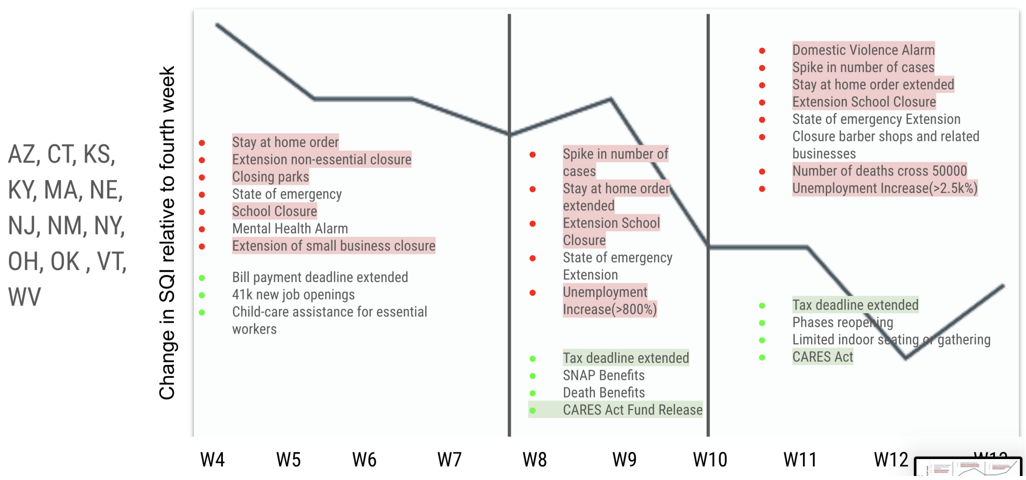

Fig. 5: Declining SQI Cluster

effect on the SQI. Further, a slight decline in SQI in the middle time interval is caused by the high increase (250%) in the unemployment rate.

A graph for declining SQI is shown in Figure 5, consisting of states: Arizona, Connecticut, Kansas, Kentucky, Massachusetts, Nevada, New Jersey, New Mexico, New York, Ohio, Oklahoma, Vermont, West Virginia. Here, in this cluster, we can see that the financial policies such as tax deadline extended, CARES Act Fund Release shows a slight improvement in the second time frame. While, the stay at home order, school closure extension, small business closure extension, high increase (800%, 2.5k%) in the unemployment rate and spike in the number of cases and deaths have a large negative impact on the SQI.

E. Part 02 Analysis (3rd October 2020 - 31st January 2021

We compared two clusters, an improving SQI cluster and a declining SQI cluster, and mapped various events, government policies, and volume of mental health and addiction entities per 1 Million tweets. A graph of improving SQI is shown in Figure 6, consisting of states: Colorado, Connecticut, Delaware, Hawaii, Illinois, Massachusetts, Maine, Minnesota, Montana, Nevada. Here, in this cluster, we can see that the number of mental health and addiction entities is decreasing with improving SQI. We also see a similar pattern as seen in the first part of the analysis. Financial events such as high household Income support, Loan relief fund and a low increase in unemployment has a more significant positive impact. A slight decline in the second time frame is caused by a High increase in the unemployment rate, a peak in the daily increase of the number of cases. An exciting set of events caused the improvement in the third and fourth intervals because of vaccine supply for frontline workers and further high vaccination rates in the cluster of states. The cluster is a major blue state, and from the news, it can be seen that blue states readily accepted vaccination and improved SQI [25].

A graph of declining SQI is shown in Figure 7, consisting of states: Alaska, Arkansas, Florida, Iowa, Idaho, Indiana, Mississippi, Rhode Island, Tennessee, Wisconsin, West Virginia. Here, in this cluster, we can see that the number of mental health and addiction entities is increasing with declining SQI. We also see a similar pattern as seen in the first part of the

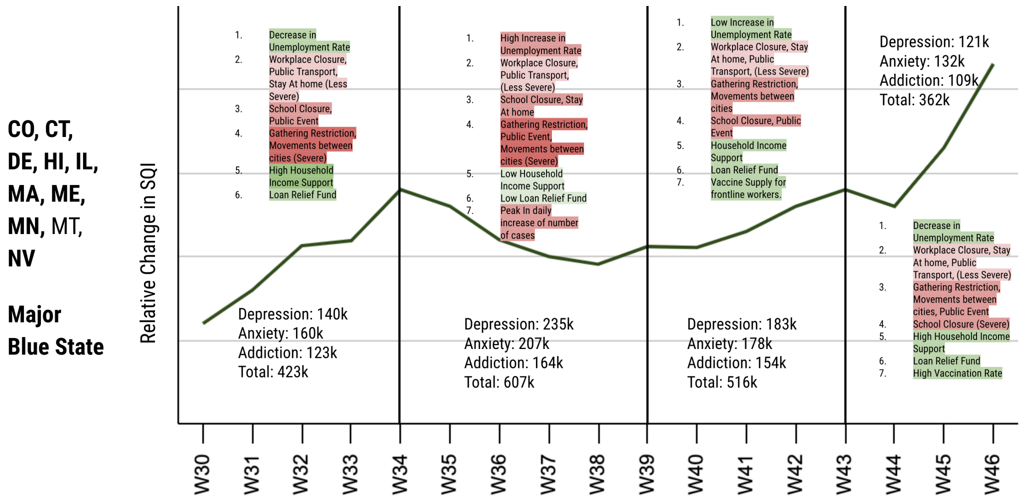

Fig. 6: Improving SQI Cluster

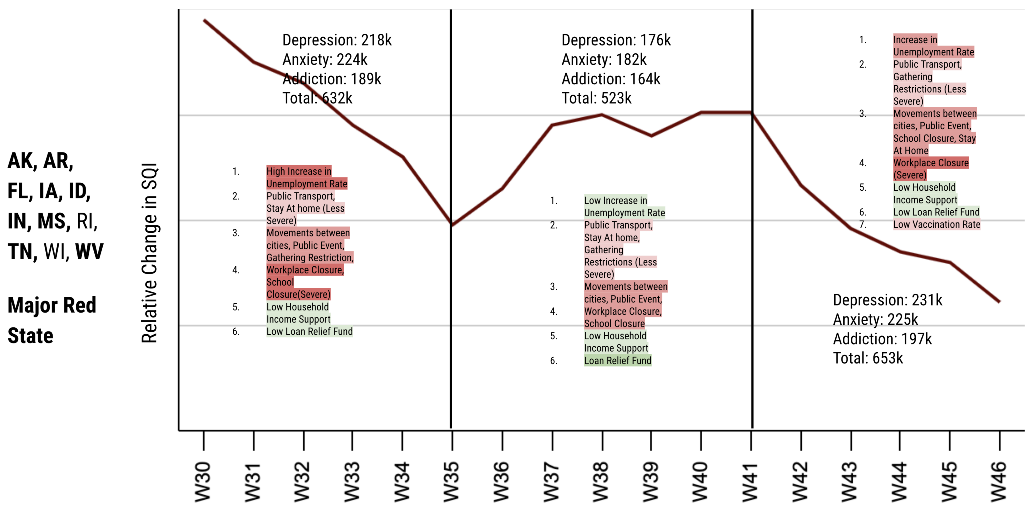

Fig. 7: Declining SQI Cluster

analysis. The financial event has a slightly positive impact which can be seen in the second time frame with a low increase in unemployment, household income support and loan relief fund. However, overall, the high unemployment rate, workplace, business closures and stay at home, movements, and gathering restrictions negatively impact the SQI. The low vaccination rate is an exciting set of events causing a decline in the third time frame. Even though the vaccine was introduced, the cluster of states had a low vaccination rate. The cluster is also a major red state, and from the news, it can be seen that red states were hesitant towards the vaccine and had protested against it [25].

VI. Conclusion

The work identifies various trust issues with the compliance of the policies and various other political and social factors affecting it. This work aims to identify and not provide solutions for the trust issues. The policymakers can use the insights from the data to understand and rectify the issues with the policy implemented at different demographics. Further, public health specialists can be prepared for the sudden hike in the queries to mental health emergency hotlines. As part of future work, identify the statistical correlation between changes in SQI ranking, unemployment rate and mental health issues.

References

[1] "Coronavirus cases:." [Online]. Available: https://www.worldometers.info/coronavirus/

[2] "Coronavirus job losses hit these states the hardest, according to record unemployment claims." [Online]. Available: https://finance.yahoo.com/news/coronavirus-hit-jobs-in-these-states-the-hardest-according-to-record-unemployment-claims-211207347.html

[3] "Public trust in federal government near historic lows for more than a decade," Sep 2020. [Online]. Available: https://www.pewresearch.org/politics/2020/09/14/americans-views-of-government-low-trust-but-some-positive-performance-ratings/pp_09-14-20_views-of-government-00-1/

[4] Q. Han, B. Zheng, M. Cristea, M. Agostini, J. J. Bélanger, B. Gützkow, J. Kreienkamp, N. P. Leander, P. Collaboration *et al.*, "Trust in government regarding covid-19 and its associations with preventive health behaviour and prosocial behaviour during the pandemic: a cross-sectional and longitudinal study," *Psychological medicine*, pp. 1–11, 2021.

[5] D. Devine, J. Gaskell, W. Jennings, and G. Stoker, "¡? covid19?¿ trust and the coronavirus pandemic: What are the consequences of and for trust? an early review of the literature," *Political Studies Review*, p. 1478929920948684, 2020.

[6] G. Rasul, “A framework for improving policy priorities in managing covid-19 challenges in developing countries,” *Frontiers in Public Health*, vol. 8, 2020.

[7] N. Joseph, P. Grover, P. K. Rao, and P. V. Ilavarasan, “Deep analyzing public conversations: Insights from twitter analytics for policy makers,” in *Conference on e-Business, e-Services and e-Society*. Springer, 2017, pp. 239–250.

[8] R. Aswani, A. K. Kar, and P. V. Ilavarasan, “Experience: managing misinformation in social media—insights for policymakers from twitter analytics,” *Journal of Data and Information Quality (JDIQ)*, vol. 12, no. 1, pp. 1–18, 2019.

[9] A. Jackson, “A crisis mental-health hotline has seen an 891% spike in calls,” Apr 2020. [Online]. Available: https://www.cnn.com/2020/04/10/us/disaster-hotline-call-increase-wellness-trnd/index.html

[10] T. N. Richards, J. Nix, S. M. Mourtgos, and I. T. Adams, “Comparing 911 and emergency hotline calls for domestic violence in seven cities: What happened when people started staying home due to covid-19?” *Criminology & Public Policy*, 2021.

[11] echen102, “Echen102/covid-19-tweetids: The repository contains an ongoing collection of tweets ids associated with the novel coronavirus covid-19 (sars-cov-2), which commenced on january 28, 2020.” [Online]. Available: https://github.com/echen102/COVID-19-TweetIDs

[12] OxCGRT, “Oxcgrt/covid-policy-tracker: Systematic dataset of covid-19 policy, from oxford university.” [Online]. Available: https://github.com/OxCGRT/covid-policy-tracker

[13] MultiState, “Covid-19 policy tracker.” [Online]. Available: https://www.multistate.us/issues/covid-19-policy-tracker

[14] . Published: Dec 02, “State covid-19 data and policy actions,” Dec 2021. [Online]. Available: https://www.kff.org/coronavirus-covid-19/issue-brief/state-covid-19-data-and-policy-actions/

[15] “Covid-19 recovery - united states department of state,” Nov 2021. [Online]. Available: https://www.state.gov/covid-19-recovery/

[16] “Diagnostic and statistical manual of mental disorders (dsm–5).” [Online]. Available: https://www.psychiatry.org/psychiatrists/practice/dsm

[17] U. Lokala, R. Daniulaityte, F. Lamy, M. Gaur, K. Thirunarayan, U. Kursuncu, and A. P. Sheth, “Dao: An ontology for substance use epidemiology on social media and dark web,” *JMIR Public Health and Surveillance*, 2020.

[18] [Online]. Available: https://www.geonames.org/ontology/documentation.html

[19] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, “Dbpedia: A nucleus for a web of open data,” in *The semantic web*. Springer, 2007, pp. 722–735.

[20] “Home feed subreddits (0).” [Online]. Available: https://www.reddit.com/subreddits/

[21] E. Kodirov, T. Xiang, and S. Gong, “Semantic autoencoder for zero-shot learning,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3174–3183.

[22] M. Gaur, U. Kursuncu, A. Alambo, A. Sheth, R. Daniulaityte, K. Thirunarayan, and J. Pathak, “” let me tell you about your mental health!” contextualized classification of reddit posts to dsm-5 for web-based intervention,” in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 753–762.

[23] K.-Y. Pan, A. A. Kok, M. Eikelenboom, M. Horsfall, F. Jörg, R. A. Luteijn, D. Rhebergen, P. van Oppen, E. J. Giltay, and B. W. Penninx, “The mental health impact of the covid-19 pandemic on people with and without depressive, anxiety, or obsessive-compulsive disorders: a longitudinal study of three dutch case-control cohorts,” *The Lancet Psychiatry*, vol. 8, no. 2, pp. 121–129, 2021.

[24] hjr5089, “Samhsa reports on covid-19 and behavioral health disparities in black and latino communities,” Jun 2020. [Online]. Available: https://www.porh.psu.edu/samhsa-reports-on-covid-19-and-behavioral-health-disparities-in-black-and-latino-communities/

[25] “The red/blue divide in covid-19 vaccination rates,” Sep 2021. [Online]. Available: https://www.kff.org/policy-watch/the-red-blue-divide-in-covid-19-vaccination-rates/