

<https://doi.org/10.46344/JBINO.2020.v09i06.06>

NEWS ADVERSELY AFFECTING THE SPREAD OF COVID-19

Chakor Pandkar & Amogh Kulkarni
Maharashtra Institute of Tecnology Pune, Maharashtra, India
Pune Institute of Computer Technology Pune, Maharashtra, India

Email : chakorpandkar@gmail.com

Abstract

Starting in China's Wuhan province, Covid-19, in a matter of few months, spread on a global level. Amongst the many factors which caused its spread, public perception is left unaccounted for. The current portrayal of the pandemic in news and social media seems to have taken a negative toll on the public sentiment, and for good reason. We hypothesize that a negative public sentiment would affect the precautions taken, consequently affecting the growth rate of covid-19. We analyze the performance of various regression models, using their R2 scores, on the growth rate of covid-19 cases with respect to the public sentiment (extracted via twitter analysis). We also extracted significant relationships between the public sentiment and the associated Covid-19 growth parameters. The regression models were most sensitive to positive sentiment based values (Near zero p-values), and not to negative sentiment-based attributes. A higher positive sentiment corresponded to a slower growth rate of the cases, while negative sentiments did not seem to affect the growth rate.

Key Words- *Machine Learning, Coronavirus, Public Perception, Sentiment Analysis, News Public Perception*

I. INTRODUCTION

The novel coronavirus, first reported to be in China's Wuhan on 31st December, 2019, has spread across the globe in the matter of a few months. Declared as a pandemic on 11th March, 2020, a quick google search gave us the counts as follows (as of 12th September, 2020) [1] :

- Confirmed Cases: 28.3M
- Recovered: 19.1M
- Deaths: 913K

A. Modes of Transmission

Covid-19 is primarily transmitted through droplets of various sizes, and via contact routes [2]. Transmission via droplets occurs when a healthy person comes in close proximity to another infected person, and therefore is at a risk of having his respiratory tract getting infected by the virus (through eyes, nose, and mouth). Transmission also occurs via infected surfaces, such as clothes, furniture, and various other points of contact. [3] It is important to note that transmission can even occur from asymptomatic people, i.e, people showing no symptoms, but who emit aerosols when they talk or breathe [4]. These particles can drift in the air for up to 3 hours. [5]

B. Public Perception

Other than the methods of transmission illustrated earlier, there one major factor which is left unaccounted for, having a pivotal role to play in Covid-19 transmission.

Not only is this factor important with respect to Covid-19, but it is important for any disease, novel or historic, which can be transmitted from person to person, and that is the perception of the public. [6]

Currently, the way the pandemic has been perceived by the public via the news, social media, and other forms of communication, has been quite negative[7]. We hypothesize that this negative sentiment (or positive as well) should affect the growth rate of the cases[8]. Also, we need to account for the gestation period of Covid-19 (We keep a track of the new cases in the next 7 and 14 days).

The perception of the public is extremely important as it can give us a psychological insight into the sentiment of the public, which should, theoretically, directly relate to the safety and precautions they take with respect to the virus [9]. Thus, we should be able to observe a pattern of increase in cases when people have an casual approach and vice versa. In this paper, we explore this hypothesis, perform our analysis on twitter covid-19 sentiment data [10] between 1st April, 2020 to 31st July, 2020. Upon having an idea about the sentiment of the public, we perform data mining to find out if this has notable effect on Covid-19 transmission.

We have acquired the twitter dataset from [11]. We compiled the sentiments of each day into the required attributes explained later in methodology. We have acquired the dataset of covid spread from [12] and performed the required pre-processing for converting it into the required attributes (explained later in methodology).

II. RELATED WORK DONE

In [7] the authors have conducted a sentiment analysis on the news headlines of over 141208 news sources to calculate the sentiment spread by the news articles regarding Covid-19. They found that most of the articles spread a negative sentiment regarding the pandemic. Similarly in [8] the authors have deduced from a nation wide poll that a negative sentiment regarding a topic in the news, invokes negative sentiment about that topic amongst its viewers. The topic in question here was countries but the same logic can be extrapolated to all the topics. Thus they form the backbone of our hypothesis that news channels all over the world are spreading a negative sentiment regarding Covid-19 through a negative coverage of the pandemic. We attempt to model a relationship between the sentiments spread by the news and the rise in cases of Covid-19 subsequently.

The importance of public perception regarding the pandemic severely impacts the spread of the virus. Illustrations and guidelines from World Health Organization (WHO) [13] refer to this by stressing the importance of RCCE or Risk Communication and Community Engagement. Not only have they outlined goals and action steps in order to achieve those

goals but they have also identified key reasons highlighting the importance of RCCE in battling pandemics and outbreaks. In an open letter the author of [14] has also highlighted the importance of public perception and trust in the battle against the pandemic. Both of these show that public perception and perceived risk of the outbreak is a driving force in the spread of the pandemic. Although they provide guidelines on how to solve this issue, they do not go in depth to actually search and objectively state the importance of these factors.

The authors of [15] have developed a mathematical model with a few assumptions modeling the change in human behaviour in the face of an outbreak using game theory. They discuss the change in the perception of risk and risk assessment in accordance to the perception towards the outbreak and how this can be modelled.

In [16] the authors have developed an agent based model using machine learning to compare the responses and their consequences of an individual tackling an epidemic versus of a collective. According to the author the process of risk assessment in a pandemic is a complex process with several important factors such as severity and information received, etc. The paper goes in depth over the varied response towards risk perception of an individual vs a collective with a clear leadership. The authors conclusively prove that the responses of the collective with a clear leadership outperform the responses of an individual.

Although these papers perform an in-depth analysis of the human behaviour and their perceived risk assessment in a pandemic, they have not calculated the correlation between the human sentiment/perception and its impact on the spread of the pandemic. We have thus used several machine learning models, to model a relation between the human sentiments (recorded via twitter analysis) and the spread of this pandemic in order to obtain a predictor at a global level.

III. METHODOLOGY

A. About the Data

Our dataset consists of various attributes related to public's sentiment with respect to covid, and the associated growth in cases observed. The public sentiment was extracted via users' tweets. Each tweet was accordingly given a sentiment score. The average number of tweets (with their associated sentiment scores) is approximately 2 million per day. More about this dataset is available at [11]. On the basis of the sentiment per day, we were able to extract the following attributes which could be potentially useful for our research:

- Mean Positive (Mean positive sentiment of all tweets for a given day)
- Mean Negative (Mean negative sentiment of all tweets for a given day)
- Percent Positive (Percent of positive tweets for a given day)
- Percent Negative (Percent of negative tweets for a given day)
- Percent Neutral (Percent of neutral tweets for a given day)
- Total Positive (Total number of positive tweets for a given day)

- Total Negative (Total number of negative tweets for a given day)
- Total Neutral (Total number of neutral tweets for a given day)
- Standard deviation (from Mean Positive)
- Standard deviation (from Mean Negative)

Note: Neutral sentiment has value 0, hence there is no need for mean/standard deviation of neutral attribute

For the data of growth of covid-19 cases, the following attributes were considered:

- New cases (on that day)
- New deaths (on that day)
- Total cases (till date)
- Total deaths (till date)
- Weekly cases (New cases over the next week from that day)
- Weekly deaths (New deaths over the next week from that day)
- Biweekly cases (New cases over the next 14 days from that day)
- Biweekly deaths (New deaths over the next 14 days from that day)

For our research, we considered the days between 1st April, 2020 and 31st July, 2020, giving us a time frame of 122 (+14 for 31st July's biweekly columns) days.

B. Data Visualization

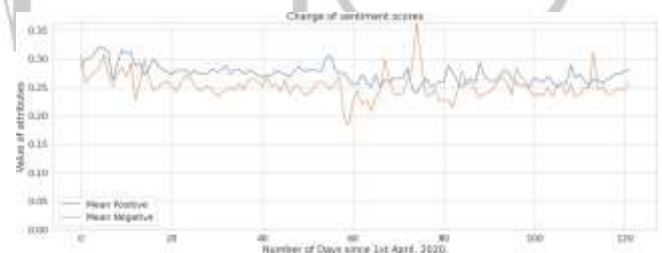


Fig. 1: Fluctuations in mean positive and mean negative values per day

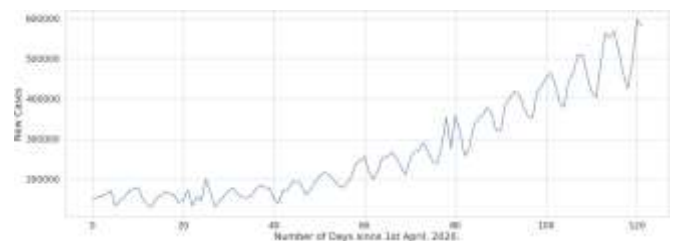


Fig. 2: New cases per day

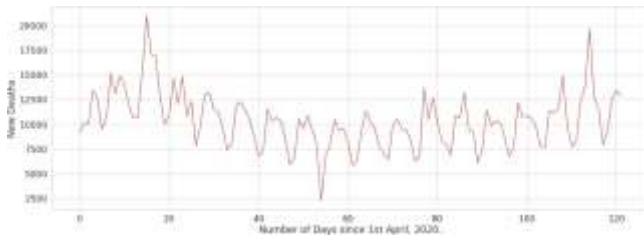


Fig. 3: New deaths per day

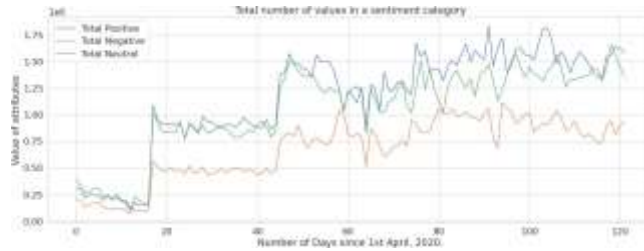


Fig. 4: Fluctuations in total positive, negative and neutral sentiments per day

Two important points to note in Fig. 4 are that:

- On 18th April, more corona specific keywords were added, leading to sharp increase in number of tweets captured.
- On 16th May, the servers were optimized, leading to significant rise in number of tweets captured per day, hence the sharp rise.

IV. EXPERIMENTAL RESULTS

A. Model Performance per Dependent Variable

Dependant Variable	MAE	MSE	RMSE	R2 Score
new_cases	0.0245571	0.0009728	0.0311891	0.2945000
new_death	0.0228613	0.0008057	0.0283857	0.0397675
total_cases	0.0249167	0.0011230	0.0335113	0.4848875
total_deaths	0.0152426	0.0004165	0.0204075	0.7001288
weekly_cases	0.0206232	0.0006577	0.0256461	0.4564516
weekly_deaths	0.0093723	0.0001240	0.0111344	0.4645634
biweekly_cases	0.0182962	0.0005612	0.0236899	0.5298348
biweekly_deaths	0.0088570	0.0001071	0.0103508	0.5007425

TABLE I: Linear Regression

Dependant Variable	MAE	MSE	RMSE	R2 Score
new_cases	0.0247331	0.0008156	0.0285584	0.4084929
new_death	0.0200989	0.0007487	0.0273630	0.1077109
total_cases	0.0263842	0.0010098	0.0317779	0.5367986
total_deaths	0.0149916	0.0003629	0.0190495	0.7387098
weekly_cases	0.0223554	0.0006363	0.0252259	0.4741158
weekly_deaths	0.0101307	0.0001535	0.0123915	0.3368345
biweekly_cases	0.0209164	0.0005646	0.0237620	0.5269677
biweekly_deaths	0.0103570	0.0001455	0.0120631	0.3218915

TABLE II: Bayesian Ridge Regression

Dependant Variable	MAE	MSE	RMSE	R2 Score
new_cases	0.0314800	0.0013815	0.0371682	-0.0019211
new_death	0.0207925	0.0008415	0.0290082	-0.0028120
total_cases	0.0380588	0.0021911	0.0468095	-0.0050463
total_deaths	0.0307048	0.0013914	0.0373017	-0.0018753
weekly_cases	0.0299753	0.0012221	0.0349580	-0.0099267
weekly_deaths	0.0132786	0.0002320	0.0152304	-0.0018300
biweekly_cases	0.0299045	0.0012013	0.0346602	-0.0064340
biweekly_deaths	0.0129829	0.0002150	0.0146623	-0.0018080

TABLE III: Lasso (Least Angle) Regression

Dependant Variable	MAE	MSE	RMSE	R2 Score
new_cases	0.0431960	0.0022317	0.0472407	-0.6185405
new_death	0.0278411	0.0011847	0.0344194	-0.4118337
total_cases	0.0445444	0.0026728	0.0516995	-0.2260041
total_deaths	0.0309738	0.0013917	0.0373058	-0.0020911
weekly_cases	0.0355284	0.0015416	0.0392637	-0.2740293
weekly_deaths	0.0154453	0.0002948	0.0171699	-0.2732349
biweekly_cases	0.0330958	0.0013923	0.0373139	-0.1664506
biweekly_deaths	0.0143137	0.0002457	0.0156755	-0.1450411

TABLE IV: Support Vector Regressor

Dependant Variable	MAE	MSE	RMSE	R2 Score
new_cases	0.0218914	0.0007478	0.0273456	0.4576672
new_death	0.0203740	0.0007105	0.0266557	0.1532448
total_cases	0.0235378	0.0009364	0.0306009	0.5704773
total_deaths	0.0143185	0.0003350	0.0183035	0.7587746
weekly_cases	0.0183640	0.0005293	0.0230056	0.5626155
weekly_deaths	0.0092797	0.0001236	0.0111158	0.4663560
biweekly_cases	0.0172936	0.0004682	0.0216386	0.6077315
biweekly_deaths	0.0095582	0.0001195	0.0109312	0.4431793

TABLE V: ARD Regressor

As per our findings, Lasso LARS and Support Vector Regression performed the worst, with negative R2 Scores for each dependent variable. Linear, Bayesian Ridge, and ARD performed better, with positive R2 scores per dependent variable. Amongst those 3, ARD performed the best in terms of average R2 scores. Hence, we concentrated our focus on ARD Regression for the next part of the analysis.

B. OLS Regression Results per Dependent Variable

Independent Var.	coef	std err	t	P> t	[0.025	0.975]
mean_positive	-0.6247	0.178	-3.502	0.001	-0.98	-0.27
per_neutral	0.2246	0.11	2.047	0.044	0.006	0.443
total_positive	1.4194	0.17	8.365	0	1.082	1.757
std_positive	222.0048	58.823	3.774	0	104.965	339.045

TABLE VI: Dependent Variable: new_cases

Independent Var.	coef	std err	t	P> t	[0.025	0.975]
perc_positive	0.2976	0.021	14.501	0	0.257	0.338
total positive	-0.3606	0.089	-4.046	0	-0.538	-0.183

TABLE VII: Dependent Variable: new_death

Independent Var.	coef	std err	t	P> t	[0.025	0.975]
mean_positive	-0.9026	0.191	-4.714	0	-1.284	-0.522
per_neutral	0.2983	0.118	2.533	0.013	0.064	0.533
total_positive	1.8487	0.182	10.15	0	1.486	2.211
std_positive	253.6134	63.142	4.017	0	127.98	379.247

TABLE VIII: Dependent Variable: total_cases

Independent Var.	coef	std err	t	P> t	[0.025	0.975]
mean_positive	-0.6514	0.162	-4.027	0	-0.973	-0.33
total_neutral	0.8809	0.214	4.12	0	0.455	1.306
total_positive	0.4671	0.232	2.015	0.047	0.006	0.928
std_positive	157.1553	39.487	3.98	0	78.574	235.737

TABLE IX: Dependent Variable: total_deaths

Independent Var.	coef	std err	t	P> t	[0.025	0.975]
mean_positive	-0.6502	0.154	-4.223	0	-0.957	-0.344
per_neutral	0.2158	0.095	2.278	0.025	0.027	0.404
total_positive	1.5	0.146	10.242	0	1.209	1.791
std_positive	240.1058	50.77	4.729	0	139.09	341.122

TABLE X: Dependent Variable: weekly_cases

Independent Var.	coef	std err	t	P> t	[0.025	0.975]
std_positive	87.7998	10.871	8.077	0	66.178	109.421

TABLE XI: Dependent Variable: weekly_deaths

Independent Var.	coef	std err	t	P> t	[0.025	0.975]
mean_positive	-0.6137	0.145	-4.238	0	-0.902	-0.326
per_neutral	0.1891	0.089	2.123	0.037	0.012	0.366
total_positive	1.5062	0.138	10.935	0	1.232	1.78
std_positive	237.8428	47.752	4.981	0	142.832	332.854

TABLE XII: Dependent Variable: biweekly_cases

Independent Var.	coef	std err	t	P> t	[0.025	0.975]
per_neutral	0.1117	0.023	4.887	0	0.066	0.157
total_positive	0.245	0.053	4.641	0	0.14	0.35
std_positive	120.5873	19.858	6.073	0	81.084	160.091

TABLE XIII: Dependent Variable: biweekly_deaths

New cases, weekly cases, biweekly cases and total cases are all affected by the mean positive sentiment, total positive tweets and standard deviation of the positive sentiments from its mean value (illustrated by a near-zero P value). They are also affected by percent neutral tweets but with a higher P value than the other factors. New deaths are affected by factors like percent positive tweets and total positive tweets received.

Total Deaths are affected by factors like mean positive sentiment, total positive tweets and standard deviation of the positive sentiments from its mean value. Total neutral tweets received also affects the total deaths.

Weekly deaths are affected by standard deviation of the positive sentiments only.

Bi-weekly deaths are affected by factors like total positive tweets, standard deviation of positive tweets and percent neutral tweets.

The percent neutral just made the cut in terms of its significance levels.

Inferences:

- A high positive standard deviation coefficient shows that it is important to the model.
- A negative coefficient for mean positive (in new cases, total cases, weekly cases and biweekly cases) shows an inverse relationship between them. From this, we can state that a higher positive sentiment was indicative of an informed risk assessment and perception of the situation, thus leading to a decrease in rate of Covid-19 growth.
- A contradiction was observed with respect to new deaths and total deaths. The polarity in mean positive coefficient indicates that the model failed to accurately account for its significance in death-related attributes. Most likely, the global death count is immune to public sentiment, but dependant on the health-care availability, medical infrastructure, etc.

V. CONCLUSION

The way the pandemic has been portrayed in the news and social media has aroused negative feelings amongst the people. The absence of negative sentiment based attributes in terms of significance values, coupled with the ubiquitous and inverse relation of positive attributes (with respect to new weekly cases), shows that the negative perception of general public did not affect the rate of growth of Covid-19 while a positive perception did. From this, we can infer that instead of spreading awareness via negativity (daily increase in cases, fatality rates, hyperbolizing the gravity of the pandemic etc), the authorities should spread awareness in a positive light (future improvements, reduced growth rates etc).

VI. FUTURE SCOPE

This research can be complimented by identifying key factors that alter public sentiment. Furthermore a more comprehensive sentiment analysis will aid in acquiring a much more accurate sentiment value. At a country level this research will help in identifying new plausible ways to tackle the pandemic by configuring the public sentiment. Also, in case of other pandemics, this model could be used for forecasting, to obtain an early estimate of the potential gravity of the situation.

VII. ACKNOWLEDGEMENTS

We thank R. Lamsal for making the twitter dataset available via IEEE Open Access [11]. We would also like to extend our gratitude towards Hannah Ritchie for making the Covid Spread Dataset openly available at [12].

VIII. DATA AVAILABILITY STATEMENT

The twitter dataset is openly available via IEEE Open Access at [11]. The Covid 19 cases dataset is openly available at [12]. The combined extracted dataset created by us is openly available at [17].

REFERENCES

- [1] World health organization, *covid timeline*. [Online]. Available: <https://www.who.int/news-room/detail/29-06-2020-covidtimeline> (visited on 07/19/2020).
- [2] *Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations*. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/331601/WHO-2019-nCoV-Sci_Brief_Transmission_modes-2020.1-eng.pdf (visited on 05/27/2020).
- [3] S. Asadi, N. Bouvier, A. S. Wexler, and W. D. Ristenpart, "The coronavirus pandemic and aerosols: Does covid-19 transmit via expiratory particles?" *Aerosol Science and Technology*, vol. 54, no. 6, pp. 635–638, 2020, PMID: 32308568. DOI: 10.1080/02786826.2020.1749229. eprint: <https://doi.org/10.1080/02786826.2020.1749229>. [Online]. Available: <https://doi.org/10.1080/02786826.2020.1749229>.
- [4] Y. Bai, L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, and M. Wang, "Presumed Asymptomatic Carrier Transmission of COVID-19," *JAMA*, vol. 323, no. 14, pp. 1406–1407, Apr. 2020, ISSN: 0098-7484. DOI: 10.1001/jama.2020.2565. eprint: https://jamanetwork.com/journals/jama/articlepdf/2762028/jama_bai_2020_id_200013.pdf. [Online]. Available: <https://doi.org/10.1001/jama.2020.2565>.
- [5] *Harvard medical school*. [Online]. Available: <https://www.health.harvard.edu/diseases-and-conditions/preventing-the-spread-of-the-coronavirus> (visited on 07/19/2020).
- [6] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, "The role of the airline transportation network in the prediction and predictability of global epidemics," *Proceedings of the National Academy of Sciences*, vol. 103, no. 7, pp. 2015–2020, Feb. 2006, ISSN: 1091-6490. DOI: 10.1073/pnas.0510525103. [Online]. Available: <http://dx.doi.org/10.1073/pnas.0510525103>.
- [7] A. Faheem, A. T. Mumtaz, S. J. Hussain, K. Aisha, and P. Mahwish, "Sentiments and emotions evoked by news headlines of coronavirus disease (covid-19) outbreak," *Humanities and Social Sciences Communications*, vol. 7, no. 23, 2020. DOI: 10.1057/s41599-020-0523-3. eprint: <https://doi.org/10.1057/s41599-020-0523-3>. [Online]. Available: <https://doi.org/10.1057/s41599-020-0523-3>.
- [8] W. Wanta, G. Golan, and C. Lee, "Agenda setting and international news: Media influence on public perceptions of foreign nations," *Journalism & Mass Communication Quarterly*, vol. 81, no. 2, pp. 364–377, 2004. DOI: 10.1177/107769900408100209. eprint: <https://doi.org/10.1177/107769900408100209>. [Online]. Available: <https://doi.org/10.1177/107769900408100209>.
- [9] F. Bagnoli, P. Liò, and L. Sguanci, "Risk perception in epidemic modeling," *Phys. Rev. E*, vol. 76, p. 061 904, 6 Dec. 2007. DOI: 10.1103/PhysRevE.76.061904. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevE.76.061904>.
- [10] V. A. Kharde and P. S. Sonawane, *Sentiment analysis of twitter data: A survey of techniques*, 2016. arXiv: 1601.06971.
- [11] R. Lamsal, *Coronavirus (covid-19) tweets dataset*, 2020. DOI: 10.21227/781w-ef42. [Online]. Available: <http://dx.doi.org/10.21227/781w-ef42>.
- [12] *Coronavirus Source Data*. [Online]. Available: <https://ourworldindata.org/coronavirus-source-data> (visited on 09/12/2020).
- [13] M. Khosravi, *Perceived Risk of COVID-19 Pandemic: The Role of Public Worry and Trust*. DOI: 10.29333/ejgm/7856. [Online]. Available: <https://doi.org/10.29333/ejgm/7856> (visited on 03/26/2020).
- [14] *Risk communication and community engagement readiness and response to coronavirus disease (COVID-19)*. [Online]. Available: <https://apps.who.int/iris/bitstream/handle/10665/331513/WHO-2019-nCoV-RCCE-2020-2-eng.pdf> (visited on 03/19/2020).
- [15] S. Zhao, Y. Kuang, C.-H. Wu, K. Bi, and D. Ben-Arieh, "Risk perception and human behaviors in epidemics," *IIEE Transactions on Healthcare Systems Engineering*, vol. 8, no. 4, pp. 315–328, 2018. DOI: 10.1080/24725579.2018.1464085. eprint: <https://doi.org/10.1080/24725579.2018.1464085>. [Online]. Available: <https://doi.org/10.1080/24725579.2018.1464085>.
- [16] A. SA, A. E-W, F. T, M. K, and M. YT, "Risk perception and behavioral change during epidemics: Comparing models of individual and collective learning," *PLoS ONE*, vol. 5, no. 1, 2020. DOI: 10.1371/journal.pone.0226483. eprint: <https://doi.org/10.1371/journal.pone.0226483>. [Online]. Available: <https://doi.org/10.1371/journal.pone.0226483>.
- [17] *Coid 19 growth and corresponding twitter sentiment analysis*. [Online]. Available: <https://github.com/amoghkulkarni5/covid-19-twitter-sentiment-dataset> (visited on 07/19/2020).