Sentiment Nowcasting during the COVID-19 Pandemic

Ioanna Miliou[1], John Pavlopoulos[1], and Panagiotis Papapetrou[1]

Stockholm University, Stockholm, Sweden
{ioanna.miliou, ioannis, panagiotis}@dsv.su.se

Abstract. In response to the COVID-19 pandemic, governments around the world are taking a wide range of measures. Previous research on COVID-19 has focused on disease spreading, epidemic curves, measures to contain it, confirmed cases, and deaths. In this work, we sought to explore another essential aspect of this pandemic, how do people feel and react to this reality and the impact on their emotional well-being. For that reason, we propose using epidemic indicators and government policy responses to estimate the sentiment, as this is expressed on Twitter. We develop a nowcasting approach that exploits the time series of epidemic indicators and the measures taken in response to the COVID-19 outbreak in the United States of America to predict the public sentiment at a daily frequency. Using machine learning models, we improve the short-term forecasting accuracy of autoregressive models, revealing the value of incorporating the additional data in the predictive models. We then provide explanations to the indicators and measures that drive the predictions for specific dates. Our work provides evidence that data about the way COVID-19 evolves along with the measures taken in response to the COVID-19 outbreak can be used effectively to improve sentiment nowcasting and gain insights into people's *current* emotional state.

Keywords: sentiment · nowcasting · COVID-19 · Twitter · measures.

1 Introduction

Epidemics of infectious diseases are triggered by factors such as changes in the ecology of a population or a novel pathogen. One such example is the outbreak of COVID-19, which resulted in a substantial burden to the world in terms of health risks and unnecessary deaths as well as financial risks and global economic turmoil. Identifying the optimal sequence of mitigation measures is always a challenge [17], with countries all over the world adopting different policies to control and limit the impact of the pandemic. Such decisions vastly rely on epidemic models (e.g., compartmental models [19]) that attempt to capture and reflect epidemic indicators, such as infection rate, recovery rate, deaths, and population mobility [2]. Recent work on reinforcement learning has also attempted to identify optimal mitigation policies [12, 13] by defining reward functions considering the impact of the pandemic on public health and the economy.

Nonetheless, the attention to public sentiment has been limited as a result of the pandemic and the mitigation measures taken [11]. Sentiment analysis concerns the classification of the intentions of a text's author (e.g., of a tweet) as positive, negative, or neutral. This is a well-known field in Natural Language Processing [23], and it is often applied on social media texts [22]. When the outcome is emotions instead of a sentiment class, the task is called textual emotion recognition [6] (also known as emotion prediction, detection, or classification). Recent works have started to explore the automated analysis of sentiments of social media posts related to the recent COVID-19 pandemic as a means to understand people's behaviors and responses during the pandemic [11, 21]. Previous research on nowcasting sentiment has employed different datasets. Either to measure consumer sentiment with the use of Google search data [5, 7] or to predict people's mood using Twitter data [14] and several other heterogeneous data (Twitter, Facebook, mood forms, mobile phone use data, and sensor data) [20].

The effects of COVID-19 contact minimization, isolation measures, lockdowns, as well as the potential fear of infection and death can have an undoubtedly long-term negative psychological and emotional impact on the population. This can, in turn, lead to severe indirect socio-economical consequences [8]. In this paper, our goal is to explore the emotional well-being of the population and identify potential factors that contribute to negative emotions related to the COVID-19 pandemic. More concretely, we propose a workflow for nowcasting negative sentiment, as expressed by Twitter, using governmental mitigation policies and epidemic indicators as exogenous variables.

Our main contributions can be summarized as follows: (1) We propose a sentiment nowcasting workflow for predicting the daily sentiment in response to the mitigation measures and epidemic indicators related to the COVID-19 pandemic; (2) We employ a sentiment extraction approach from tweets using a transformer-based, multi-lingual, masked language model called XLM-R; (3) Our workflow supports both statistical as well as machine learning models. For the latter, it also provides explanations for the predictions in the form of local model agnostic explainable features, using LIME; (4) Our empirical evaluation on data including tweets, mitigation measures, and epidemic indicators, obtained over two periods during the development of the pandemic suggests that mitigation measures and epidemic indicators can potentially function as factors for predicting negative public sentiment.

2 Sentiment nowcasting

We propose a workflow for estimating the negative sentiment value related to COVID-19 one day ahead of the latest ground-truth value by taking advantage of exogenous variables, such as epidemic indicators and mitigation measures. Let T define a set of tweets written in natural language. The first step of our workflow is to convert T to a time series of sentiment values by employing a function $g(\cdot)$. Let $Y = y_1, \ldots, y_t$ denote the time series of t real-valued sentiment observations, with each $y_i \in \mathbb{R}$. We additionally consider a set of exogenous variables $\boldsymbol{X} = \{\mathcal{X}^j\}$

that can occur concurrently with Y. Each $\mathcal{X}^j$ comprises a set of variables that together correspond to some common exogenous factor that can contribute to the estimation of Y. The main goal of this paper is to define a function $f(\cdot)$ that predicts the next observation $y_{t+1} = f(Y, \boldsymbol{X})$ by taking into account both the historical observations of Y, as well as the sets of exogenous variables up to time t. In our setup, Y models the degree of negative sentiment per day, which is extracted as an aggregate value from a set of tweets that are filtered based on language and location. Moreover, we employ two sets of exogenous variables. The first ($\mathcal{X}^1$) contains 20 indicators and four indices that correspond to government mitigation policies for COVID-19, while the second ($\mathcal{X}^2$) contains 55 epidemic metrics related to the development of COVID-19.

Sentiment extraction Each tweet was annotated regarding sentiment by XLM-R, a multilingual, Transformer-based model [3]. We fine-tuned XLM-R to extract sentiment for a tweet as a valence score from zero (very negative) to one (very positive), and we binarized that score by using a threshold (see Section 3.1). Then, y_t is the fraction of the negative tweets out of the filtered tweets of day t.

Data Smoothing We decided to smooth the sentiment data to eliminate noise and random fluctuations. This allows important patterns to more clearly stand out and is intended to ignore one-time outliers. We choose to apply a Trailing Moving Average. The value at time t is calculated as the average of the raw observations over a time window of length $w = 3$ ending at time t.

Statistical Models Models handling time series are used in order to predict future values of indices by extracting relevant information from historical data. Traditional time series models are based on various mathematical approaches, such as autoregression. For this study, we apply the models of **Autoregression**, **Exponential Smoothing**, **ARIMA**, and **ARIMAX**.

Machine Learning Models We used regression models to assess whether the inclusion of mitigation measures and epidemic indicators can improve the accuracy of the classical methods. Regression analysis is a form of predictive technique that models the relationship between a dependent (target) and one or more independent variables (predictor). In our case, the target is the negative sentiment expressed on Twitter and the predictors are the epidemic indicators and the mitigation measures. For this study, we apply **Linear Regression**, **Ridge Regression**, **Lasso Regression**, **Random Forest**, and **eXtreme Gradient Boosting (XGBoost)**. For each model, the best hyperparameters are selected in each training phase by Grid Search and 10-fold Cross-validation.

3 Empirical evaluation

3.1 Data description

Sentiment evaluation data To evaluate the performance of sentiment extraction through XML-R, we used the SemEval-2018 Affect in Tweets sentiment dataset (V-reg), which considers sentiment as a score from zero (very negative) to one (very positive) [15]. More specifically, scores below 42.9% indicate the negative sentiment class, scores above 61% indicate the positive sentiment class, and scores in between indicate the neutral sentiment class. The dataset consists of 2,567 tweets that were annotated by 175 annotators (49,856 annotations reported in total), and it is already split into 1,181 tweets for training, 449 for validation, and 937 for testing.

Twitter COVID-19 The tweets that were used in our study were obtained through the Twitter Streaming API. Considering we are interested in capturing the sentiment during the COVID-19 pandemic, we filtered the tweets that comprise COVID-19 related keywords. Our data spans two chronological periods. The first period is from 3/11/2020 to 17/12/2020, but unfortunately, we have a few missing dates for a total of 32 days of available data. The second period is from 20/4/2021 to 14/5/2021, for a total of 25 days. We have more than 13 million tweets and an average of 179,000 tweets per day. We filter the tweets based on their location to include only tweets from the USA, and we obtain almost 654,000 tweets. The fraction of negative tweets (threshold of 42.9%; see the above paragraph) is 317,000.

Mitigation measures The Oxford COVID-19 Government Response Tracker (OxCGRT) was designed to systematically collect information on different common policy responses taken by governments in response to the pandemic [9]. It contains data from 186 countries on various policies, including school closures, stay-at-home orders, economic support for households, and vaccination. The data is publicly available [1], and more concretely comprises 20 indicators of government responses that can be grouped into three categories: (1) Containment and closure policies (indicators C1-C8), such as school closures and restrictions in movement, (2) Economic policies (indicators E1-E4), such as income support to citizens or provision of foreign aid, and (3) Health system policies (indicators H1-H8), such as the COVID-19 testing regime, emergency investments into healthcare, and most recently, vaccination policies. The data from these 20 indicators is aggregated into a set of four indices: (1) Overall government response index; (2) Containment and health index; (3) Economic support index; (4) Stringency index.

Epidemic indicators We additionally use the COVID-19 dataset maintained by Our World in Data [18]. The data is updated daily throughout the COVID-19 pandemic covering 226 countries and territories on 55 metrics, including (1) confirmed cases and deaths, (2) hospitalizations and intensive care unit (ICU) admissions, (3) tests and positivity data, (4) vaccination data, (5) other variables of interest.We should note that due to the long reporting chain of new cases and deaths, the daily reported number does not necessarily represent the actual number on each day. For that reason, negative values in cases and deaths may appear if a country corrects previously overestimated historical data.

3.2 Setup

We choose to study the United States of America (USA), an English-speaking country, as it is better represented in the Twitter dataset. We construct the statistical and machine learning models to produce the sentiment predictions. Initially, we split our datasets into training and test sets (85%-15%). The training data is used to estimate and generate the models' parameters, and the test data is used to calculate the accuracy of the models. However, at every step of the training, we update the training set with the latest historical value, and the models are retrained (i.e., we employ dynamic training). Thus, the models are updated with the latest information available to include any fluctuation in the sentiment, indicating an increase in COVID-19 cases, a new measure taken, etc.

At each step, we obtain a new predicted value for the sentiment. Once the training is completed, we have our predictions according to the initial test set's length. Then, we evaluate the accuracy of the predictions with respect to the initial test set that contains the actual sentiment values. We consider standard performance indicators to evaluate the performance of the predictive models: the Pearson Correlation, the Mean Absolute Percentage Error (MAPE), and the Root Mean Square Error (RMSE) [4, 10].

3.3 Results

Sentiment extraction We used XLM-R, which achieves a root mean square error (RMSE) of 0.015 and a mean absolute percentage error (MAPE) of 0.261 on the test set of the SemEval-2018 V-reg sentiment dataset. The high predictive power of the model, reflected by the low error, makes it a suitable candidate for our sentiment annotation task. We note that our data comprises more than a million tweets, making human annotation impossible. On the same evaluation dataset, NLTK's Sentiment Intensity Estimation baseline model achieves a much worse RMSE (0.053) and MAPE (0.529) score.[1]

Statistical models Table 1 presents the performance indicators for the statistical models that use only the endogenous variable, i.e., the sentiment. We notice that the three models, Autoregression, Exponential Smoothing, and `ARIMA` perform very similarly with respect to the three metrics, and they are able to capture somehow the way the sentiment evolves in time, but not very accurately.

Table 1. Performance indicators of the statistical models.

Model	Pearson	MAPE	RMSE
Autoregression	0.620	1.502	0.009
Exp Smoothing	0.626	1.560	0.009
`ARIMA`	0.610	1.580	0.009

We conclude that the sentiment time series itself is not sufficient to predict the future sentiment. For that reason, we attempt to estimate the future sentiment by including different sets of exogenous variables into the `ARIMAX` model. Table 2 presents the performance indicators for the `ARIMAX` model that makes use of

[1] http://www.nltk.org/howto/sentiment.html

Table 2. Performance indicators of the `ARIMAX` model. We use the sentiment variable along with the independent variables referring to the COVID-19 indicators (COV), mitigation measures (ME), and a combination of all (COV&ME).

Model	Pearson			MAPE			RMSE		
	COV	**ME**	**COV&ME**	**COV**	**ME**	**COV&ME**	**COV**	**ME**	**COV&ME**
`ARIMAX`	0.258	0.563	0.051	3.750	2.021	4.484	0.021	0.011	0.025

Table 3. Performance indicators of the machine learning models. We make use of the independent variables referring to the COVID-19 indicators (COV), mitigation measures (ME) and a combination of all (COV&ME).

Model	Pearson			MAPE			RMSE		
	COV	**ME**	**COV&ME**	**COV**	**ME**	**COV&ME**	**COV**	**ME**	**COV&ME**
Linear	0.232	**0.598**	0.520	4.581	2.017	**1.763**	0.027	0.014	**0.011**
Ridge	**0.616**	0.610	0.612	2.318	**1.917**	2.323	**0.013**	**0.013**	0.014
Lasso	**0.716**	0.355	0.714	**1.623**	2.185	1.721	**0.009**	0.013	**0.009**
RF	0.868	0.666	**0.870**	1.718	**1.455**	1.570	**0.009**	**0.009**	**0.009**
XGBoost	0.666	0.618	**0.892**	1.427	1.451	**1.082**	0.008	0.009	**0.006**

the sentiment variable along with the independent variables referring to the COVID-19 indicators (COV), mitigation measures (ME), and a combination of all (COV&ME). We see that not only the exogenous variables do not improve the performance, but in the case of COV&ME, where the number of features is very high, the model fails to predict the future sentiment. We explain such a result due to the nature of the model that assumes a linear relationship between the target variable and the various features. In this case, the model is incapable of selecting only the relevant features, resulting in the inclusion of noisy signals.

Machine Learning models At this point, we choose to explore the possibility of estimating the sentiment more accurately with the machine learning models that use the different sets of exogenous variables without incorporating any autoregressive behavior. Table 3 presents the performance indicators for the machine learning models. We test the performance of Linear, Ridge, and Lasso Regressions, as well as Random Forest (RF) and XGBoost. We test the models with the use of a) the COVID-19 indicators (COV), b) the mitigation measures (ME), c) a combination of the COVID-19 indicators and mitigation measures (COV&ME). Overall, the XGBoost outperforms the other prediction models in terms of Pearson correlation, MAPE, and RMSE. It outperforms all the statistical models, as well as the other machine learning models. More specifically, we observe that the best results are obtained when we make use of both the epidemic indicators and mitigation measures (COV&ME), with a Pearson correlation of 0.892, MAPE of 1.082, and RMSE of 0.006.

Fig. 1 presents the ground-truth sentiment time series along with the predictions from the XGBoost model with the COVID-19 indicators and mitigation measures (COV&ME). We observe that the predictions are able to monitor the evolution of the sentiment accurately over time. Moreover, Table 3 and Fig. 1 reflect the added value of using the epidemic indicators and measure data over

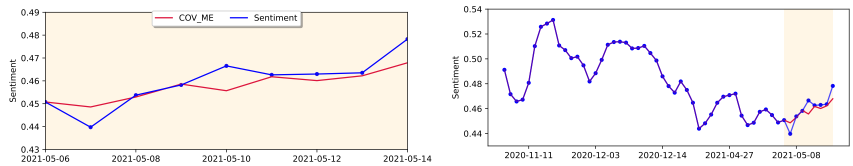

Fig. 1. Ground-truth sentiment time series along with the predictions from the XGBoost model with the COVID-19 indicators and mitigation measures (COV&ME). The left plot zooms in the predictions time frame of the right plot.

the historical autoregressive and Exponential Smoothing approaches. Forecasts obtained with XGBoost are significantly more accurate for COV&ME compared to the statistical models.

Feature importance In general, AI models make it difficult, even for the experts, to explain the rationale of their conclusions. For that reason, we consider it crucial to provide understandable results, not only to verify their correctness and quality but, above all, to explain what drives the sentiment of the people during the COVID-19 pandemic period. A benefit of using XGBoost is that it is easy to retrieve importance scores for each feature. Generally, importance provides a score indicating how useful or valuable each feature was in constructing the boosted decision trees within the model. The more an attribute is used to make key decisions with decision trees, the higher its relative importance. We notice in Fig. 1 that the biggest error is for the prediction of 2021-05-10. For that reason, we choose to analyze this particular record to explain what drives the prediction on this day.

Fig. 2 (left) shows the 20 most important features (importance on the x-axis) that drive the prediction of 2021-05-10, as calculated from XGBoost with the epidemic indicators and mitigation measures (COV&ME). The most important feature is "stringency_index" with an importance of 0.461, which records the strictness of "lockdown style" policies that primarily restrict people's behavior. We have several COVID-19 indicators that score between the most important features related to tests, cases, deaths, and ICU patients. Additionally, we have the "GovernmentResponseIndex" which records the government's response throughout the outbreak, and the "H7_Vaccination policy" indicator that records the vaccination policy of the health system.

Local explanations Feature importance measures rarely provide insight into the average direction that a feature affects the response function. They state the magnitude of a feature's relationship with the response compared to other features used in the model. We cannot know specifically the influence of each factor for a single observation. We hence decided to use LIME, which stands for Local Interpretable Model-agnostic Explanations [16] to help us understand individually what features and how they influence the sentiment of each day. LIME is a novel explanation technique that explains the prediction of any classifier or regressor in an interpretable and faithful manner by approximating it locally

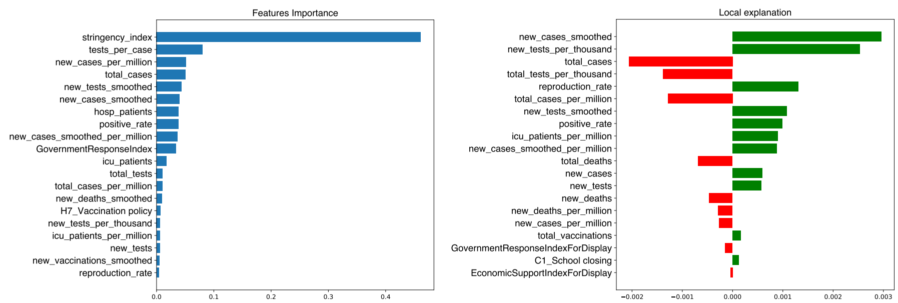

Fig. 2. Local explanation of the 20 most important features that drive the prediction of the 2021-05-10, as calculated by XGBoost (left) and with LIME for XGBoost (right) with the COVID-19 indicators and mitigation measures (COV&ME).

with an interpretable model. LIME supports explanations for tabular models, text classifiers, and image classifiers.

Fig. 2 (right) provides the local explanation for the most important features that drive the prediction of 2021-05-10 and their relative strength. Each feature is then color-coded to indicate the relative increase or decrease in the prediction probability, i.e., whether the feature supports or increases the prediction value (Green) or it has a negative effect or decreases the prediction value (Red), respectively. For example, "new_cases_smoothed" is the most important feature with a weight of 0.003, and "new_tests_per_thousand" the second most important with a weight of 0.0025, both of green color, which indicates that they increase the value of the prediction. On the contrary, "total_cases" and "total_tests_per_thousand" are red, indicating a decrease in the prediction value.

Comparing the most important features obtained from the XGBoost model and LIME, we see that the two approaches have 15 out of 20 features in common. That is a very good indicator of the goodness of this result. To further verify if those 15 features are indeed the most relevant for our predictions, we test the accuracy of the predictions with the `ARIMAX` model that suffered greatly from the increased number of features. In Table 4 we report the results from an `ARIMAX` model that uses all the features and an `ARIMAX` model that uses only these 15 (`ARIMAX`-15). Comparing the two models, we observe that feature filtering improves the performance of `ARIMAX` in terms of the three metrics. We should note that `ARIMAX` does not outperform the simpler statistical models. Our objective, however, is to emphasize the improvement in its performance with the filtered features and not to suggest it as a better model. XGBoost remains as the model with the best performance when using both the epidemic indicators and mitigation measures (COV&ME).

4 Conclusions

In this paper, we highlighted the limitations of earlier work in considering the direct emotional and psychological impact of the mitigation measures taken in

Table 4. Performance indicators of the `ARIMAX` model with all the features and of the `ARIMAX` model with only the 15 (`ARIMAX`-15).

Model	Pearson	MAPE	RMSE
`ARIMAX`-COV&ME	0.051	4.484	0.025
`ARIMAX`-15	0.604	1.774	0.009

response to the COVID-19 pandemic. We hence proposed a workflow for identifying potential exogenous factors that can be used for the task of negative sentiment nowcasting, employing both statistical and machine learning models. Our results suggest that machine learning models, such as XGBoost, can substantially improve sentiment nowcasting compared to standard autoregressive models. Directions for future work include the exploration of multivariate statistical models, such as VARMAX, as well as RNN- and Transformer-based architectures. We will also explore more extensive feature selection technique as an assistive preprocessing step to statistical models, and the use of data collected over a more extended period, e.g., over the whole pandemic.

Acknowledgements

The work of IM and PP has been supported in part by the Digital Futures EXTREMUM project titled "Explainable and Ethical Machine Learning for Knowledge Discovery from Medical Data Sources". The work of PP has also been supported in part by the Vinnova project titled "Artificial Intelligence for Mitigation of Pandemics".

References

1. Blavatnik School of Government, University of Oxford: Covid-19 government response tracker. https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker#data (2021), (Online; accessed June 2021)
2. Cobey, S.: Modeling infectious disease dynamics. Science **368**(6492), 713–714 (2020)
3. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V.: Unsupervised cross-lingual representation learning at scale. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 8440–8451. Online (Jul 2020)
4. De Myttenaere, A., Golden, B., Le Grand, B., Rossi, F.: Mean absolute percentage error for regression models. Neurocomputing **192**, 38–48 (2016)
5. Della Penna, N., Huang, H., et al.: Constructing consumer sentiment index for us using google searches. Tech. rep. (2010)
6. Deng, J., Ren, F.: A survey of textual emotion recognition and its challenges. IEEE Transactions on Affective Computing (2021)
7. Duncan, B., Elkan, C.: Nowcasting with numerous candidate predictors. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 370–385. Springer (2014)

8. Gismero-Gonzalez, E., Bermejo-Toro, L., Cagigal, V., Roldan, A., Martinez-Beltran, M.J., Halty, L.: Emotional impact of covid-19 lockdown among the spanish population. Frontiers in Psychology **11**, 3634 (2020)
9. Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., Tatlow, H.: A global panel database of pandemic policies (oxford covid-19 government response tracker). Nature Human Behaviour **5**(4), 529–538 (2021)
10. James, G., Witten, D., Hastie, T., Tibshirani, R.: An introduction to statistical learning, vol. 112. Springer (2013)
11. Kabir, Y., Madria, S.: EMOCOV: Machine learning for emotion detection, analysis and visualization using COVID-19 tweets. Online Social Networks and Media **23**, 100135 (2021)
12. Kompella, V., Capobianco, R., Jong, S., Browne, J., Fox, S., Meyers, L., Wurman, P., Stone, P.: Reinforcement learning for optimization of covid-19 mitigation policies. arXiv preprint arXiv:2010.10560 (2020)
13. Kwak, G.H., Ling, L., Hui, P.: Deep reinforcement learning approaches for global public health strategies for covid-19 pandemic. Plos one **16**(5), e0251550 (2021)
14. Lansdall-Welfare, T., Lampos, V., Cristianini, N.: Nowcasting the mood of the nation. Significance **9**(4), 26–28 (2012)
15. Mohammad, S., Bravo-Marquez, F., Salameh, M., Kiritchenko, S.: SemEval-2018 task 1: Affect in tweets. In: Proceedings of The 12th International Workshop on Semantic Evaluation. pp. 1–17. Association for Computational Linguistics, New Orleans, Louisiana (Jun 2018)
16. Ribeiro, M.T., Singh, S., Guestrin, C.: “why should i trust you?” explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. pp. 1135–1144 (2016)
17. Richard, Q., Alizon, S., Choisy, M., Sofonea, M.T., Djidjou-Demasse, R.: Age-structured non-pharmaceutical interventions for optimal control of covid-19 epidemic. PLOS Computational Biology **17**(3), 1–25 (03 2021), https://doi.org/10.1371/journal.pcbi.1008776
18. Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., Macdonald, B., Giattino, C., Appel, C., Rodés-Guirao, L., Roser, M.: Coronavirus pandemic (covid-19). Our World in Data (2020)
19. Tolles, J., Luong, T.: Modeling epidemics with compartmental models. Jama **323**(24), 2515–2516 (2020)
20. Tsakalidis, A., Liakata, M., Damoulas, T., Jellinek, B., Guo, W., Cristea, A.: Combining heterogeneous user generated data to sense well-being. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. pp. 3007–3018 (2016)
21. Yang, Q., Alamro, H., Albaradei, S., Salhi, A., Lv, X., Ma, C., Alshehri, M., Jaber, I., Tifratene, F., Wang, W., et al.: Senwave: Monitoring the global sentiments under the covid-19 pandemic. JMIR public health and surveillance **6**, e19447 (2020). https://doi.org/doi:10.2196/19447
22. Yue, L., Chen, W., Li, X., Zuo, W., Yin, M.: A survey of sentiment analysis in social media. Knowledge and Information Systems **60**(2), 617–663 (2019)
23. Zhang, L., Wang, S., Liu, B.: Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **8**(4), e1253 (2018)