In the beginning of 2020, a new viral outbreak originating from China established itself as one of the world’s biggest threats. In December 2019 a viral pneumonia of unknown aetiology was identified in Wuhan, Hubei, China (H. Lu et al., 2020). Later, on the 7th of January, the Chinese Centre for Disease Control and Prevention identified the causative agent of these viral pneumonias as a new strain of coronavirus, which was subsequently named by the World Health Organization (WHO) SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) (World Health Organization, 11 February 2020). The WHO has also named the disease caused by SARS-CoV-2 COVID-19. SARS-CoV-2 is a betacoronavirus belonging to the *Coronavirinae* subfamily, part of the Coronaviridae family (Sohrabi et al., 2020). Its genome presents 8 conserved SARS-CoV amino acids responsible for interaction with human angiotensin-converting enzyme 2 (Fehr & Perlman, 2015). SARS-CoV is a previously identified coronavirus strain which caused the 2003 SARS outbreak (R. Lu et al., 2020). COVID-19 is an acute resolved respiratory disease that can lead to death, with a reported 2 % fatality rate (Bassetti et al., 2020; Chan et al., 2020; Huang et al., 2020). Symptoms present themselves usually after an incubation period of approximately 5.2 days (Rothan & Byrareddy, 2020). The person-to-person transmission of COVID-19 occurs through direct contact or through droplets derived from coughing or sneezing of infected patients (Rothan & Byrareddy, 2020).

On the 30th of January 2020 the WHO declared COVID-19 a Public Health Emergency of International Concern (Sohrabi et al., 2020; World Health Organization, 11 February 2020). After cases in multiple countries such as Italy, France, Germany, Spain, South Korea, Iran, USA, and evidence of local person-to-person transmission of the disease, on the 13th of March 2020 the WHO declared COVID-19 a pandemic (World Health Organization, 13 March 2020). The classification of COVID-19 as a pandemic resulted in the implementation of multiple mitigation measures by the affected countries, in order to control the outbreak. This is especially important in the first phases of the outbreak as SARS-CoV-2 has a reported average R0 of 3.28 (Liu et al., 2020). Usually as the epidemic progresses the reproductive number of a virus declines steadily until it reaches a value lower than 1 and decays (Anderson et al., 2020). This is due to lack of susceptible individuals or because of the effect of implementation of multiple mitigation measures. This is supported by simulation studies on COVID-19 transmission using SEIR models (Fang et al., 2020). The main objectives of mitigation measures are minimising morbidity and mortality, avoiding or shifting in time epidemic peaks to avoid overwhelming the health-care system and to flatten the epidemic curve (Anderson et al., 2020).

As COVID-19 has evolved to a global pandemic discussion about it has been continuing through the past year and on numerous platforms. Probably the largest amount of discussing about the pandemic has happened on social media, especially by the general public (Pastor, 2020). Due its enormous usage social media is considered one of the global centres of big data (DeNardis & Hackl, 2015). The amount of information and the length of time spent on social media by users has been shown to have an influence on the users’ behaviour (Alamoodi et al., 2021). This is influence has an effect not only on the users themselves but is also essential to governmental decision making as it is a direct catalogue of public opinion (Chung et al., 2015). One of the ways of extracting information from social media data is sentiment analysis as it can contribute to the understanding of human emotions through the users’ engagement with social media (Ji et al., 2016). A sentiment based approach to analysing the COVID-19 pandemic has already shown good results in both prediction and inference of the general public’s opinion on both mitigation measures and the pandemic as a whole (Aljameel et al., 2021; Alomari et al., 2021; Gilardi et al., 2021; Wang et al., 2020). Not many studies have tried to measure the direct effect that social media has on a government’s ability to take difficult decisions such as those needed in the fight against COVID-19. A Swiss study on this issue has though found that social media does influence the ability and ease of discussion about mitigation measures by politicians (Gilardi et al., 2021).

The pressures from both the daily scientific data about the pandemic’s course and the general public’s opinion must have an influence on governmental decisions and policy making, but the effect that both have on the strictness of decisions remains unclear.

**Plan of Research**

In this project,

sentiment (neutral, pos, neg).

challenge is that human behaviour does not follow one set of specific rules but rather tends to change from one community to another and from one person to another. In answer to this the sentiment analysis algorithms basically work through dictionaries and give a polarity to each token (see tokenization) based on the word’s usage in language.

Alessandro: I agree with this; that’s why possibly including more countries could be interesting (i.e English speaking countries, Italy since I speak Italian and could translate the searched words in the algorithm, and Switzerlan/Germany?)

**Data**

MICHAEL PART1

**Research Questions**

MICHAEL PART2

**Data processing and analysis**

1. Clean the dataset and filter for the three countries
2. Exploratory data analysis (EDA) to conduct summary statistics and check for correlation
3. Sentiment analysis & modification of sentiment analysis for other languages (Italian and German):
4. Tokenization (segregation into words)
5. Cleaning (removing the special characters)
6. Removing Stop words (preposition, auxiliary verbs, etc.)
7. Classification of words (+1: positive, -1 negative, 0: neutral)
8. Apply supervised algorithm for classification (train model with word or lexicons, and test on the analysis statement)
9. Calculate sentiment of statement (look at polarity)
10. Categorization of mitigation measures (response variable) into strict, medium, and light
11. Predictive modelling of mitigation measure based on sentiment and pandemic course
12. Importantly the response and predictors are polytomous variables (>2 possible categories)
13. Possible machine learning models:
14. Logistic regression
15. Artificial neural network
16. If time: temporal analysis of repone changes (i.e., first wave, second wave) using the same predictors

Alamoodi, A. H., Zaidan, B. B., Zaidan, A. A., Albahri, O. S., Mohammed, K. I., Malik, R. Q., Almahdi, E. M., Chyad, M. A., Tareq, Z., Albahri, A. S., Hameed, H., & Alaa, M. (2021). Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Systems with Applications*, *167*, 114155. https://doi.org/10.1016/j.eswa.2020.114155

Aljameel, S. S., Alabbad, D. A., Alzahrani, N. A., Alqarni, S. M., Alamoudi, F. A., Babili, L. M., Aljaafary, S. K., & Alshamrani, F. M. (2021). A Sentiment Analysis Approach to Predict an Individual’s Awareness of the Precautionary Procedures to Prevent COVID-19 Outbreaks in Saudi Arabia. *International Journal of Environmental Research and Public Health*, *18*(1), 218.

Alomari, E., Katib, I., Albeshri, A., & Mehmood, R. (2021). COVID-19: Detecting Government Pandemic Measures and Public Concerns from Twitter Arabic Data Using Distributed Machine Learning. *International Journal of Environmental Research and Public Health*, *18*(1), 282.

Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). *How will country-based mitigation measures influence the course of the COVID-19 epidemic ?* *2019*(20), 931–934. https://doi.org/10.1016/S0140-6736(20)30567-5

Bassetti, M., Vena, A., & Giacobbe, D. R. (2020). The novel Chinese coronavirus (2019-nCoV) infections: Challenges for fighting the storm. *European Journal of Clinical Investigation*, *50*(3), e13209. https://doi.org/10.1111/eci.13209

Chan, J. F.-W., Yuan, S., Kok, K.-H., To, K. K.-W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C.-Y., Poon, R. W.-S., Tsoi, H.-W., Lo, S. K.-F., Chan, K.-H., Poon, V. K.-M., Chan, W.-M., Ip, J. D., Cai, J.-P., Cheng, V. C.-C., Chen, H., … Yuen, K.-Y. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet*, *395*(10223), 514–523. https://doi.org/https://doi.org/10.1016/S0140-6736(20)30154-9

Chung, W., He, S., & Zeng, D. (2015). *eMood: Modeling emotion for social media analytics on Ebola disease outbreak*.

DeNardis, L., & Hackl, A. M. (2015). Internet governance by social media platforms. *Telecommunications Policy*, *39*(9), 761–770. https://doi.org/https://doi.org/10.1016/j.telpol.2015.04.003

Fang, Y., Nie, Y., & Penny, M. (2020). Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis. *Journal of Medical Virology*, *92*(6), 645–659. https://doi.org/10.1002/jmv.25750

Fehr, A. R., & Perlman, S. (2015). *Coronaviruses: An Overview of Their Replication and Pathogenesis BT - Coronaviruses: Methods and Protocols* (H. J. Maier, E. Bickerton, & P. Britton (eds.); pp. 1–23). Springer New York. https://doi.org/10.1007/978-1-4939-2438-7\_1

Gilardi, F., Gessler, T., Kubli, M., & Müller, S. (2021). *Social Media and Policy Responses to the COVID-19 Pandemic in Switzerland*.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, *395*(10223), 497–506. https://doi.org/https://doi.org/10.1016/S0140-6736(20)30183-5

Ji, X., Chun, S. A., & Geller, J. (2016). Knowledge-Based Tweet Classification for Disease Sentiment Monitoring. In W. Pedrycz & S.-M. Chen (Eds.), *Sentiment Analysis and Ontology Engineering: An Environment of Computational Intelligence* (pp. 425–454). Springer International Publishing. https://doi.org/10.1007/978-3-319-30319-2\_17

Liu, Y., Gayle, A. A., Wilder-smith, A., & Rocklöv, J. (2020). *The reproductive number of COVID-19 is higher compared to SARS coronavirus*. *Figure 1*, 1–4. https://doi.org/10.1093/jtm/taaa021

Lu, H., Stratton, C. W., & Tang, Y.-W. (2020). Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *Journal of Medical Virology*, *92*(4), 401–402. https://doi.org/10.1002/jmv.25678

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., … Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, *395*(10224), 565–574. https://doi.org/https://doi.org/10.1016/S0140-6736(20)30251-8

Pastor, C. K. (2020). Sentiment analysis of Filipinos and effects of extreme community quarantine due to coronavirus (Covid-19) pandemic. *Available at SSRN 3574385*.

Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease ( COVID-19 ) outbreak. *Journal of Autoimmunity*, *109*(February), 102433. https://doi.org/10.1016/j.jaut.2020.102433

Sohrabi, C., Alsafi, Z., Neill, N. O., Khan, M., Kerwan, A., Al-jabir, A., Iosifidis, C., & Agha, R. (2020). World Health Organization declares global emergency : A review of the 2019 novel coronavirus ( COVID-19 ). *International Journal of Surgery*, *76*(February), 71–76. https://doi.org/10.1016/j.ijsu.2020.02.034

Wang, S., Schraagen, M., Sang, E. T. K., & Dastani, M. (2020). *Public Sentiment on Governmental COVID-19 Measures in Dutch Social Media*.