COVID-19 Spreading Rate Prediction and its Impact on Mental Health using Machine Learning

Sudip Podder, Anik Alvi, Minhajuddin Ahmed

Problem Statement:

Covid-19 has caused tremendous damage to man-kind. Governments have failed to address how to handle and confine this virus during the first wave of Covid-19. So, predicting how it is spreading in real time is very important to warn people of possible danger. Pandemic has also shown us the importance of public health in our society. Especially mental health has become an important issue during lockdown. Lifestyle before and during pandemic has changed significantly. Government and public policy makers need to address these issues since many people feel depressed and may attempt sucide. Analyzing these problems will not only help people but also help transform our society for better development during the era of COVID-19.

Motivation:

Mental health is a condition which is often neglected, and can be as dangerous as leading a person to commit suicide. In this era of social media, people choose various platforms to share their feelings. Our aim is to analyze if Covid-19 spreading rate and rise in emotional depression has any common correlation around the world. Such analysis will include social media text data and various social and environmental parameters using machine learning algorithms. We believe this analysis will be helpful for public health policy makers and the government to take accurate initiatives in minimizing these issues.

Data Collection:

For social media data, we have collected tweets from twitter in two stages. First, we collected 50,000 tweets from around the world for January and February which will give an overview of the situation before the first real impact of Covid-19. We have collected another 50,000 tweets for the duration of May to September 2020 when the Covid cases were at its peak.

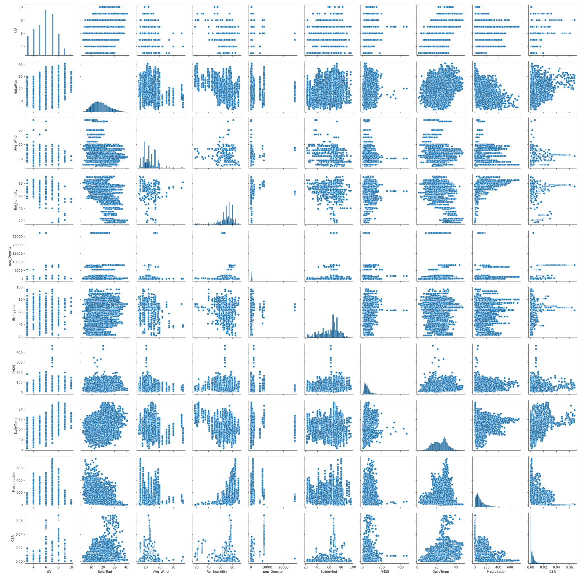
To analyze the correlation between Covid spread and depression in society, we have collected different environmental parameters daily between the time period of May to September like UV radiation, Solar radiation, average daily wind speed, relative humidity, daily average temperature, precipitation and particulate matter (PM2.5) levels which may impact spread of Covid-19. In addition, we collected some social parameters like population density, stringency index and population of 57 different cities from around the world to get an overall scenario around the world. The cities considered are as follows:

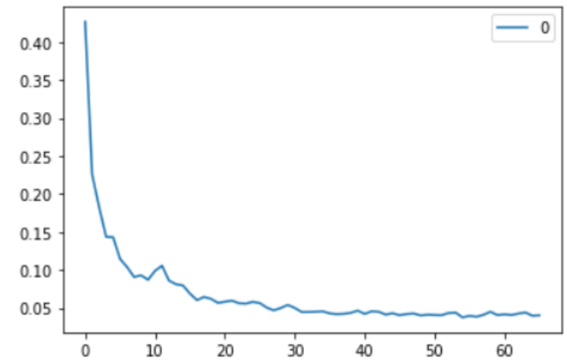
Dubai, Islamabad, Sasolburg, Delhi, Dhaka, Kuala Lumpur, Ho Chi Minh City, San Agustin Aguascalientes, South Jakarta, Istanbul, Nur-Sultan, San Jose, São Paulo, Colombo, Central Singapore, Beijing, Bangkok, Tel Aviv, Edmonton, Los Angeles City, Oklahoma City, Atlanta, New Orleans City, Tallahassee, Portland, New York City, Melbourne, Kumanovo, Las Vegas, Seoul, Vancouver, Tempe Maricopa Phoenix, Auckland, Taipei, St. John's, Toronto, Dnipropetrovs'k, Kollarova Trnava, Oslo, Lancaster, Central Hong Kong, London, Madrid, Warsaw, Brussels, Austin, Amsterdam, Vienna, Copenhagen, Valjevo, Litomerice, Paris, Budapest, Milan Lombardi, Helsinki, Thessaloniki, and Tokyo.

Since, we want to predict the Covid Spreading Rate (CSR), we have calculated the target variable using formula:

CSR = Cumulative number of Cases / Total population

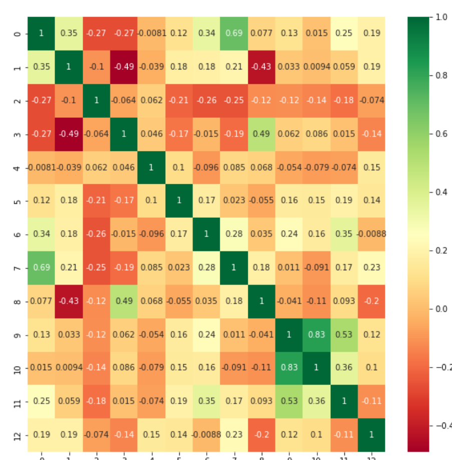
Moreover, to visualize the relationship between independent features and CSR, pair-wise plots have been presented below.





Feature Analysis:

To check how much the features are correlated to each other, we have plotted the correlation matrix as presented below.



From the correlation matrix, it can be seen that COVID cases and death rates are highly correlated to each other. As a result, the COVID-19 death rate feature can be dropped for better performance.

Methodology:

We have adopted the deep neural network architecture to perform regression analysis on the COVID-19 data. To train the regression neural network model, we split the data into training (70%) and testing (30%). We have used feature scaling so that all features have the same value range. In order to find the optimal parameters of the multi-layer neural network model, we have used grid search based cross validation technique. The parameters that were optimized were activation function, optimizer, sizes of the hidden layers, learning rate, and regularization.

Based on grid search, the following parameters were best found to optimize. Best Hyperparameters: {'activation': 'logistic', 'alpha': 0.0001, 'hidden_layer_sizes': [22, 20, 10, 5], 'learning_rate_init': 0.05, 'solver': 'adam'}

Training based on these parameters, the model converges after 60 iterations with a Mean Absolute Error of 0.1867 for training dataset and 0.2467 for testing dataset. Following diagram shows the convergence of the training.

Reference:

1. Worldometer, March 03, 2021, Accessed on: March 03, 2022. [Online]. Available: https://www.worldometers.info/.

2. Aqicn, March 03, 2021, Accessed on: March 03, 2022. [Online]. Available: https://aqicn.org/

3. AirNow, March 03, 2021, Accessed on: March 03, 2022. [Online]. Available: https://www.airnow.gov/

4. Weatherbase, March 03, 2021, Accessed on: March 03, 2022. [Online]. Available: https://www.weatherbase.com/weather/countryall.php3

5. John Hopkins University COVID-19 Data Repository, March 03, 2021, Accessed on: March 03, 2022. [Online]. Available: https://github.com/CSSEGISandData/COVID-19

6. Data World Bank, March 03, 2021, Accessed on: March 03, 2022. [Online]. Available: https://data.worldbank.org/indicator/EN.POP.DNST

7. Statistics Canada, March 03, 2021, Accessed on: March 03, 2022. [Online]. Available: https://www12.statcan.gc.ca/

8. Australia Bureau of Statistics, March 03, 2021, Accessed on: March 03, 2022. [Online]. Available: https://www.abs.gov.au/

9. World Population Review, March 03, 2021, Accessed on: March 03, 2022. [Online]. Available: https://worldpopulationreview.com/country-rankings/countries-by-density

10. Beijing Municipal Bureau of Statistics, March 03, 2021, Accessed on: March 03, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Beijing#cite_note-bulletin2018-4

11. Oxford University Government Response Stringency Index, March 03, 2021, Accessed on: March 03, 2022. [Online]. Available: https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker