Transmission, Incubation and Environmental Stability:

Analysis of Submissions

Devon-Eric Lee

University of Hawaii Maui College

**Demographics & observation for pandemic escalation: Aman Kumar**

In this submission to the Kaggle challenge the Roche Data Science Coalition utilized data sets to determine which populations are at risk of contracting COVID-19. Beginning with importing essential libraries and utilizing what we already know in such the elderly are more at risk of contracting the virus. They read Novel-Corona-Virus-2019-dataset managed by SRK into the notebook, which holds the information about the cumulative case counts of COVID-19 Across the world. The datasets mentioned in this challenge takes data from Worldometer which possess the similar figures for the age-group wise distribution of COVID-19 Cases. The figures mentioned there also highlights that people associated with an already existing COPD's or medical ailments have a higher risk of getting into a COVID-19 infection. Aman Kumar then created the dataset CovCSD - COVID-19 Countries Statistical Dataset and uncovered a collection of COVID-19 datasets from Kaggle. To conduct an analysis of the spread of COVID-19 they plotted a running map for observing the spread of the COVID-19 confirmed cases. From observations form Choropleth graphs, it starts in China at the epicenter with the initial cases reported in Australia, U.S., and Canada. Then it displays Europe as the epicenter and European countries having the 2nd-4th highest cases. In addition, they used bar graphs to analyze the spread of deaths from COVID-19 as well as confirmed recoveries. They analysis of the graphs displays on March 17th, 2020, 56 Days post the first confirmed case of COVID-19, the Global Count of confirmed covid-19 cases crossed 200k mark. Then within 7 days, on 24th March 2020, the Global confirmed case count reached beyond 400k mark. The graphs displayed that it took three days from March 24th, 2020 to March 27th, 2020, for global confirmed case count to reach 600k mark, and within the next two days there was more than 200 thousand confirmed cases. There was a total of 20.55% of recoveries out of the total number of cases by April 6th.

**Four possible ways to fight COVID-19: Yuan**

In this submission it was concluded that the four possible ways to fight COVID-19 are to have a strong medical response, excellent national health awareness and take a large number of nucleic acid testing, using time differences to effective use of medical resources, and mandatory isolation of everyone. It is displayed in this submission that with a strong medical response adequate medical resources, the mortality rate will remain around one percent. If enough ventilator supply is maintained, the lethality of COVID-19 to infected persons will remain at a level like that of influenza. With good national health habits, it can delay the arrival of the peak number of people diagnosed with the epidemic, and flatten the curve of the number of infections, reducing the impact of a large number of short-term new cases on the medical system. This submission also shows how time differences in the effective use of medical resources aids in effectively combatting COVID-19 by using personal protective equipment (PPE) across the country, and medical institutions across the country must use high-level PPE to avoid potential nosocomial infections. The final solution in this submission in fighting COVID-19 is isolation where it states the right to life is always the most human right. In this submission an analysis of the number of infections in each country examined. Countries were divided into four different categories according to growth from Type A, B, C, and D. In addition to utilizing an Analysis of percentage data was gathered from national infection rate, mortality and cure rate, and observing how many times the data on March 27th coincided with the data on March 20th. A logical data model was also utilized to examine the medical capabilities of countries from the general level, perfect medical level, strong national support, and dangerous situation amongst countries. This submission also used a World Bank Indicator analysis to examine each countries ability to combat COVID-19.

**Building a CORD19 Research Engine: Dwight Gunning**

In this submission a notebook was used to build a CORD Research Engine on top of a BM25 search index, along with a full document Similarity Index by storing the 786-dimension Specter Vectors in an Annoy index. The index complements each other where BM25 will be utilized for search and Specter Vectors provide accuracy in similarity matching. The indexes are complemented an intuitive user interface with both an interactive search bar as well as a natural language like Python API. The submission covers the code for this Kernel to provide insights on how to fight the SARS-COV2 pandemic, from what is known about transmission, incubation, and environmental stability, COVID-19 risk factors, virus, genetics, and evolution, vaccines and therapeutics, what has been published about medical care/ethical and social science issues/sharing and intersectoral collaboration, etc. This submission displays the steps needed to use the kernel, starting with Installing the dgunning/cord19 Library, then explaining the research papers class is a container for the metadata, and the BM25 search index. Then load\_metadata - Load the metadata.csv file, leading to cleaning the metadata. Next is creating the BM25 search index, leading to creating document vectors where one can create a document vector for each json record. And then adding a document similarity index which is based on Annoy, Annoy which is very simple and super-fast, and will return the mst similar items to a given query. The following steps is loading research papers, searching research papers, selecting research papers, selecting individual research papers, exploratory analysis and charts, identifying which clusters are COVID related, creating a BM25 index, listing of the tokens in each string by using gensim or NLTK to tokenize each abstract into a list of tokens. Then using a code that includes a function called parallel which applies a function over a collection, which is a modification of a function by the same name from the fastai v2 repository, with the main changes being using tqdm progress bars, and ensuring the output list is sorted in the same order as the input list. Next is loading JSON then implementing the specter vector API. Finally is to get the embedding, plot the embeddings, and then plotting the multiple embeddings.

References

aestheteaman01. (2020, April 24). Demographics & observation for pandemic escalation. Retrieved from <https://www.kaggle.com/aestheteaman01/demographics-observation-for-pandemic-escalation>

Dgunning. (2020, May 2). CORD Research Engine - Search and Similarity. Retrieved from <https://www.kaggle.com/dgunning/cord-research-engine-search-and-similarity>

Yuanso. (2020, April 17). Four possible ways to fight COVID-19. Retrieved from <https://www.kaggle.com/yuanso/four-possible-ways-to-fight-covid-19>