PROPOSAL OF IMPROVED VACCINE DISTRIBUTION PARADIGM

DATA DRIVERS
AKAR GUPTA (S3808546)
ANURAG DINESH KARMARKAR (S3829070)
CHINMAY PRADEEP KARANGUTKAR (S3827733)
SIDDHARTH KUMAR (S3816208)

CONTENTS:

INTRODUCTION AND BACKGROUND:

The advent of the industrial revolution ushered in a revolution in the healthcare sector, drastically increasing the average human life expectancy. This resulted in an increase from an average of 46 years in the 1900s to the current average of 76 years [4].

The rapid advancement in research, development and production of new vaccine and medication has established a new baseline form disease control. Notwithstanding the fiscal responsibility for a given drug's/vaccine's research and development, a handful of countries monopolize the spearheading of drug/vaccine production and distribution across the globe [5].

In the current battle against Covid-19, a global pandemic the likes of which has not been seen in the last 100 years, there is a precarious requirement of an ethical and revamped vaccine research, development and distribution system that allows access to lifesaving vaccines to the global mases without any impediment.

Presently, certain developed countries have made bilateral contracts with pharma companies, resulting in the prioritization of vaccine delivery based on their national interests. Countries with higher spending power can afford investing in multiple vaccine research, while countries with lower GDP can only invest after a vaccine has shown some promising results.

For instance, countries like the USA, Canada and the UK have invested heavily in research and development of the COVID vaccines by various pharma companies; in return the pharma companies with successful vaccine developments are contractually obligated to prioritize the delivery of said vaccine to these investor countries. These bilateral deals do not consider the impact of the pandemic on other countries down in the investment queue, due to their lesser spending power and underdeveloped healthcare systems [1].

As the result of this approach, there is a stockpiling of these much-needed vaccines by few countries, where the said vaccine may not even be approved for usage, while people in other countries suffer because of a lack of its accessibility. This disproportionate distribution of vaccine is not helping anyone so to address this we have come up with a solution.

We propose a decentralized Machine Learning model for restructuring the distribution of life saving vaccines throughout the globe. We will be looking at how our model will counter the ethical concerns as well as any other concern arising because of the proposed Machine Learning (ML) model. Our model will be focussed on developing a system which will optimize and enhance the distribution of essential vaccines and drugs during a pandemic which may occur in the foreseeable future.

SIGNIFICANCE:

In the early stages of any given epidemic/pandemic, the primary need is to curb the spread of the pathogen. During this effort, the production and distribution of vaccines and drugs gain the spotlight. Due to the unforeseen circumstances caused by Covid-19, the political leaders and leaders in healthcare industry came together for development of Covid-19 vaccinations, using the existing methodology of new vaccine production and distribution. This reliance on the old ways has put under immense stress on the system and the cracks in the system are quite evident [2].

As part of the existing method of new and rapid vaccine development, one of the most common implemented techniques were bilateral deals between the investing countries and the pharma companies. Under these deals the pharma companies would prioritize shipment of newly developed vaccines to the countries with the highest investment. While these deals worked in the past, such as in case of H1N1 epidemic where the spread was rapid and yet sporadic thus effectively contained, the rampant spread of Coronavirus across large geographies has proven these deals to be unethical. As a result, of these deals some countries ended up with excess vaccine supplies, while the population in other countries suffered due to inaccessibility to vaccinations.

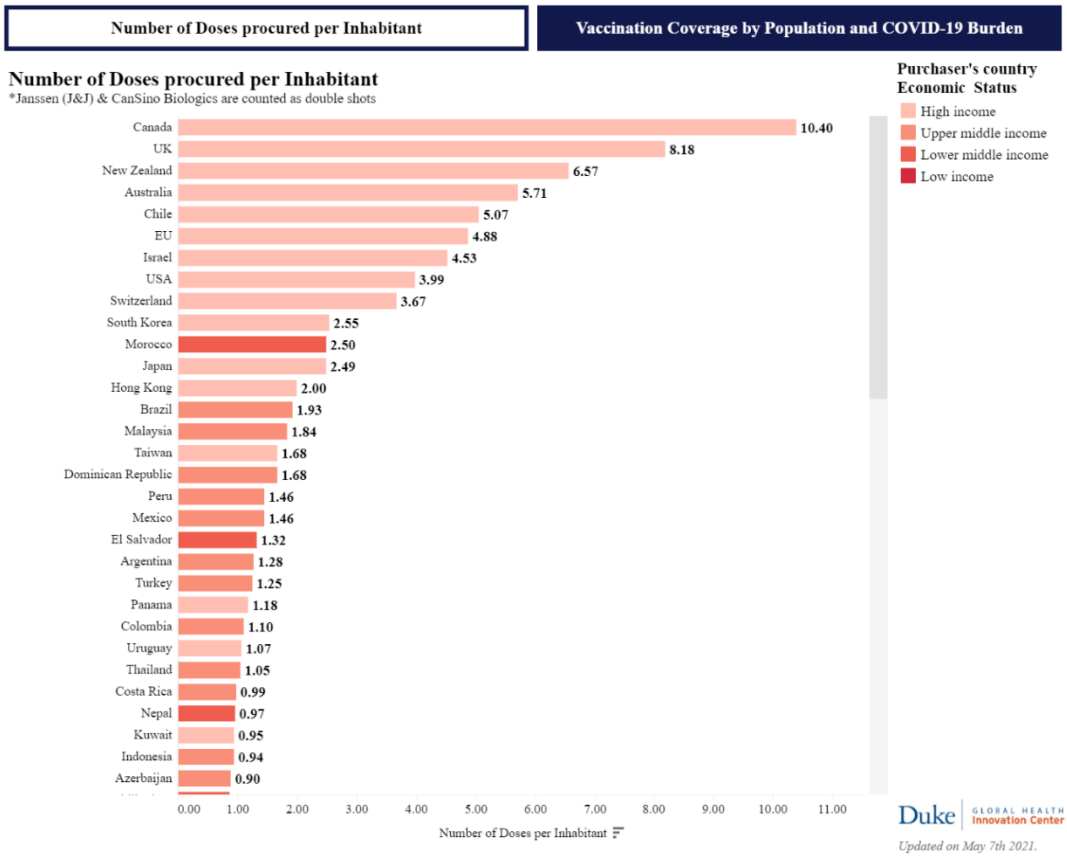
This is where COVAX comes into picture, a body established by the World Health Organization (WHO). COVAX can be considered as a supporting body helping collaborating countries access medical supplies and establish multilateral deals with pharma companies to deal with Covid-19. The major pharma companies in this domain are Johnson & Johnson, Pfizer, Moderna, AstraZeneca, Sinovac, Novavax. Currently, the pharma companies are still taking contracts from various nations in the presence of COVAX. The pharma companies directly deal with nations regarding the supply of vaccines and COVAX stands as an intermediator that monitors the COVID vaccine supply chain [2].

As a result of it, many financially powerful countries, with a higher spending power, have a hold on the vaccine supply chain as they pre-book specific amount of vaccine doses. In 2019, US contributed 1.2 billion dollars to AstraZeneca, a British Drug Manufacturing Company for boosting the COVID vaccine research in returns 300 million Covid-19 vaccine doses. Thus, it created a problematic scenario for countries with lesser spending power, as these countries fell short of investment in vaccines development resulting insufficient vaccines, failing to effectively respond to this pandemic [1]. Apart from USA, many other countries carried out the same procedures to stockpile as many doses as they can, depriving developing countries struggling for vaccines.

Many factors are working against the effectiveness of the COVAX initiative, causing significant impact to those countries reliant on the COVAX interventions. One such factor is the delayed establishment of the initiative - in the time lapsed to create the initiative, bilateral deals were set in motion, thus allowing financially sound countries to monopolize the vaccines supply, and subsequently hoarding it. Another factor impacting the initiative is the disproportionate support from developed nations in terms of vaccine stock redistribution, financial aid for enhancing the supply systems or vaccine development, capitalizing on patent rights instead of allowing free use of vaccine patents to ramp up the production and so on. This is categorically called out as a catastrophic moral failure, which our proposed solutions aim to resolve. The

issue of inequality raised when the younger and healthy peoples in economically stable nations were vaccinated before the endangered peoples in deprived countries [13].

The below graph shows the exact count of all nations covering vaccines with respect to their economic stability.



Country-wise vaccine consumption with respect to Economy [3]

Hence, this issue can be solved by an ML model that will supervise the entire vaccine supply chain. This model will be arbitrated by initiative like COVAX, which would help convey the decision-making process to the respective stake holders. The outcomes of this model will result in ensuring transparency in the supply chain methodologies and WHO can act immediately if they detect any vaccine shortage for any country.

PROPOSED IDEA/SOLUTION:

As discussed above we can see that the current infrastructure of vaccine development and distribution is not ideal. While the first come first serve approach, via bilateral deal, seems justified and works well for the developed nation in short run, in long run it puts the whole world population at risk, especially in case of rapidly mutating virus likes of Novel Coronavirus. Even with the establishment of entity like COVAX, which curtail to negotiate with pharma companies on bilateral deals has failed to server its purpose, with the major investor companies opting for bi-lateral deals [6].

The ML model we propose as a solution is a decentralized open machine learning model, which considers various parameters such as recovery rate, spread rate, population density, etc. along with the bilateral deals made by the various countries. Based on these parameters, this model would suggest additional vaccine recipients, with adjusted shipment quantities for the respective countries.

This kind of ML model would help redirect some of the vaccines to the countries hit worst by the pandemic, while still maintaining a steady stream of vaccine supply to the countries with bilateral deals. The model will also be trained to factor in stockpiling of approved and unapproved vaccines by countries and suggest redistribution of stockpiled vaccines to countries with urgent need of the same.

From the data published by the 'Duke Global Health Innovation Center', we see a clear trend of high-income countries such as Canada and England, respectively stockpiling up to 10.4 and 8.18 times the number of vaccines required to inoculate their complete population. Using our proposed ML model, we hope to solve this kind of lopsided vaccine distribution.

Considering the criticality of revamping vaccine supply, its paramount to have a ML model with a high degree of accuracy and precision to direct this crucial task.
This high degree of accuracy and precision can only be achieved via complex ML techniques, such as neural-network or gradient boosting. The inner workings of these techniques are inherently complex to understand and downright impossible to precisely convey to the stakeholder, virtually making them a black box.

This lack of transparency of the model's decision-making process makes it less trustworthy, which would deter the countries from collaborating on the proposed solution. To circumvent this problem, we would be implementing an interpretation layer on the proposed ML model. The interpretation layer would represent the intrinsic decision-making process and help understand the relationship between the input and output of the complex model.

The proposed model is designed around the F.A.T.E principle of AI/ML model development. The ML model by design would be easily interpretable, allowing the respective stakeholders to monitor and understand the decision-making process, and thus making the model **transparent** and **accountable**. Since, the model is decentralized, the vaccine distribution priority would be decided upon the ground reality, thus making it **fair** to all the stakeholders involved and would help chart out an **ethical** vaccine distribution. The transparency and impartiality of ML model would help address the current fragility of vaccine development and distribution practices by holding the pharma companies and the participating countries accountable.

METHODOLOGY:

The proposed model aims to predict the future infection rate and death rate of Covid-19 in each country and based on these predicted rates, assign a priority index to each country. This priority index would be a normalized quantity i.e., represented between 0 to 1, and would be calculated by summation of the weighted predicted infection and death rate.

The dataset used is Covid-19 datasheet published by 'Our World Data' [9]. The dataset published have a plethora of features (columns/data points) which are ***iso code, continent, location, date, total cases, new cases, new cases smoothed, total deaths, new deaths, new deaths smoothed, total cases per million, new cases per million, new cases smoothed per million, total deaths per million, new deaths per million, new deaths smoothed per million, reproduction rate, ICU patients, ICU patients per million, hospital patients, hospital patients per million, weekly ICU admissions, weekly ICU admissions per million, weekly hospital admissions, weekly hospital admissions per million, total tests, new tests, total tests per thousand, new tests per thousand, new tests smoothed, new tests smoothed per thousand, positive rate, tests per case, tests units, total vaccinations, people vaccinated, people fully vaccinated, new vaccinations, new vaccinations smoothed, total vaccinations per hundred, people vaccinated per hundred, people fully vaccinated per hundred, new vaccinations smoothed per million, stringency index, population, population density, median age, aged 65 older, aged 70 older, GDP per capita, extreme poverty, cardiovascular death rate, diabetes prevalence, female smokers, male smokers, handwashing facilities, hospital beds per thousand, life expectancy, human development index.***

The complete modelling and prediction workflow can be broken down in the following four steps:

Step 1: Feature Selection –

For a given ML model with increase in number of input features, the chances of overfitting and computational time increases. Hence, we need to opt for the process of feature selection. Under this step, we would use supervised feature selection techniques, as we need to consider the target variables. We gravitate towards 'Chi-Square' filter-based feature selection process, as it helps identify crucial features by assigning them a correlation score. The features with higher correlation score are selected as relevant features [10][11]. Also, by comparing the chi-square value of various selected features to the two target features, we can assign respective weights to the target features.

Step 2: Neural Network Implementation –

In this step, we design a deep neural network (DNN) model using the identified features as the input parameters. The dataset would be split into training, testing and validation dataset. The model hyperparameter tuning would be done by training and testing the model on training and validation dataset. The performance of the model would be evaluated based upon the mean squared loss function.

$$SE(y\_, y) = \frac{\sum_{i=1}^{n} (y - y\_)^2}{n}$$

Once we achieve a model with high accuracy and precision, and the least amount of over/under fitting, the model will be tested upon the testing dataset, which would provide an evaluation of the model performance in predicting new infection and death rate. Based on the reference paper, 'Prediction of the COVID-19 epidemic trends based on SEIR and AI models' [12], used to model this solution. We theorise that a neural network with 2 hidden layers would be most efficient.

Step 3: Decision Tree Implementation –

The previous step would generate a list of countries with their respective priority index. Now we need to decide upon the list of recipient countries with the updated distribution. To attain this, we are implementing an easy-to-understand Decision Tree machine learning model that would consider the target population under a country's vaccination policy, the optimal age range suggested by WHO for a certain vaccination policy and the priority index attained by the above model. Based on these input parameters the model would suggest the distribution quantity of the said vaccine.

Step 4: Interpretation Layer –

To attain this interpretability, we can implement either of the following depending upon the final model -

- **Surrogate Models:** These kinds of interpretation models work by training on the input datasheet and model prediction, and provide an interpretable model serving as the global summary.

- **LIME Technique** [8] : This kind of interpretation layer technique builds individual surrogates' model on the local prediction region, hence providing a much granular interpretation of the model [7]

Based on the performance of the DNN model and the level of understanding required we can select either of the above techniques.

Breaking down the above methodology based on F.A.T.E principle, given that the neural network model has input data curated through open, reliable, and diverse data sources, the likelihood of data biases is less, hence the model ensures that the decision-making process is fairer.

Being a NN model, the understanding of the decision-making process of model is quite low. Implementing an interpretability layer helps better understand the process, making the model accountable and transparent. While calculating the priority index the model does not consider the investment by a particular country. It evaluates the parameters used to determine the extent of coronavirus impact, making the model relatively ethical.

ETHICAL CONSIDERATIONS:

The proposed methodology covers some basic FATE analysis & ethical considerations. The proposed model can be considered fair as it tries to establish a clear balance between the countries, regardless of their investment. The countries that have developed the vaccine or have the supply of vaccines in abundance can become the suppliers for the countries that cannot meet the vaccine demand by creating a priority list. The economic standing of the countries will not be of importance while setting the priority of the supply. The decisions of setting the priority and the number of doses received by countries will be suggested to the group of pharma companies by our proposed model, by considering the parameters which we have identified in our model. It means that the pharma companies would be accountable for these system decisions. Transparency of the model is no longer an issue, as the model will be visible to all the collaborating countries in the world.

The model aims to be ethical as it ensures that people in all the countries of the world should have access to a COVID19 vaccine. There are some scenarios wherein ethical violations may occur in the model. Since our solution is a neural network model, it is inherently data hungry making it susceptible to biases. A collaborative organization like COVAX/WHO which consists of people of various political affiliations, which can be exploited by providing corrupt data which may lead to a biased priority list. This would result in violation of the transparency & fairness principles of our model, compromising the proposed solution. To avoid the above-mentioned issue, new legislation should be proposed to ensure data integrity.

Currently our model is working on self-reported data by various countries - this itself proposes an ethical concern as countries can incorrectly report the metrics based on which the impact of the pandemic is being calculated. If incorrect data is being fed to our neural network, it will result in a biased priority index which would result in an unethical vaccine distribution.

RESEARCH MATERIAL:

The research material used was focussed on the pre-existing model for vaccine manufacturing and distribution. It highlights the drawbacks of the current model and various factors which have been affected by these drawbacks. The data sources such as 'Our World in Data' and the 'Global Alliance for Vaccines and Immunizations' (GAVI) are considered while building the initial model [2][9]. The GAVI is a strong alliance of countries in the world who monitor the various systems to make vaccines and drugs available worldwide.

'Our World in Data' is an open-source data curation platform where data from multiple reliable and peer reviewed sources are available. The data can also be considered from other reliable sources further in developing the system. Research related to the current scenario of the COVID19 pandemic was instrumental in highlighting the various drawbacks of the current system in place [1][3][4][6][13].

Some of the research articles provided insights on various concepts in Data Science and Machine Learning[7] [8] [10] [11]. The model proposed in this report is a product of a large amount of research in the fields of healthcare, supply chain, data science and machine learning. Our proposal of a model using deep neural network is based on the research done by Data Scientists at 'Plos' [12].

REFERENCE:

1. Vox, 2021. *How rich countries are making the pandemic last longer*. [video] Available at: <https://www.youtube.com/watch?v=2ty2Jos2Woc> [Accessed 16 May 2021].
2. Gavi.org. 2020. *COVAX explained*. [online] Available at: <https://www.gavi.org/vaccineswork/covax-explained> [Accessed 16 May 2021].
3. Launchandscalefaster.org. 2020. *Vaccine Procurement | Launch and Scale Speedometer*. [online] Available at: <https://launchandscalefaster.org/covid-19/vaccineprocurement> [Accessed 16 May 2021].
4. Roser, M., Ortiz-Ospina, E. and Ritchie, H., 2020. *Life Expectancy*. [online] Our World in Data. Available at: <https://ourworldindata.org/life-expectancy> [Accessed 17 May 2021].
5. Workman, D., 2020. *Drugs and Medicine Exports by Country*. [online] Worldstopexports.com. Available at: <https://www.worldstopexports.com/drugs-medicine-exports-country/> [Accessed 18 May 2021].
6. BBC News. 2021. *Covax: How will Covid vaccines be shared around the world?*. [online] Available at: <https://www.bbc.com/news/world-55795297> [Accessed 18 May 2021].
7. Hulstaert, L., 2019. *Machine learning interpretability techniques*. [online] Medium. Available at: <https://towardsdatascience.com/machine-learning-interpretability-techniques-662c723454f3> [Accessed 18 May 2021].
8. Hulstaert, L., 2018. *Understanding model predictions with LIME*. [online] Medium. Available at: <https://towardsdatascience.com/understanding-model-predictions-with-lime-a582fdff3a3b> [Accessed 18 May 2021].
9. Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., Macdonald, B., Giattino, C., Appel, C., Rodés-Guirao, L. and Roser, M., 2021. *Coronavirus Pandemic (COVID-19)*. [online] Our World in Data. Available at: <https://github.com/owid/covid-19-data/tree/master/public/data> [Accessed 18 May 2021].
10. Brownlee, J., 2019. *How to Choose a Feature Selection Method For Machine Learning*. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/> [Accessed 18 May 2021].
11. Agarwal, R., 2019. *The 5 Feature Selection Algorithms every Data Scientist should know*. [online] Towards Data Science. Available at: <https://towardsdatascience.com/the-5-feature-selection-algorithms-every-data-scientist-need-to-know-3a6b566efd2> [Accessed 18 May 2021].
12. Feng, S., Feng, Z., Lingng, C., Chang, C. and Feng, Z., 2021. . [online] *Plos One*. Available at: < https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245101#sec007 > [Accessed 18 May 2021].
13. BBC News. 2021. *Covid vaccine: WHO warns of 'catastrophic moral failure'*. [online] Available at: <https://www.bbc.com/news/world-55709428> [Accessed 18 May 2021].