Assessment of Risk Estimates and Fatalities Involved with Covid-19

Narayana Darapaneni
Director – AIML
Great Learning/Northwestern
University Illinois, USA
darapaneni@gmail.com

Abhay Singh
Student – AIML
Great Learning
Hyderabad, India
abhayps@gmail.com

Anwesh Reddy Paduri
Data Scientist - AIML
Great Learning
Hyderabad, India
anwesh@greatlearning.in

Jayasurya M
Student – AIML
Great Learning
Hyderabad, India
mjayasurya20@gmail.com

Mukesh Kumar
Student – AIML
Great Learning
Hyderabad, India
muksjb@gmail.com

Krishna Chaitanya
Student – AIML
Great Learning
Hyderabad, India
krishna.kc69@gmail.com

Sumanth S
Student – AIML
Great Learning
Hyderabad, India
suramsumanth@gmail.com

2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) | 978-1-6654-4067-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/IEMTRONICS52119.2021.9422659

Abstract-- **The objective of this work is to understand COVID-19 spread and lethality across the world along with the factors affecting them. For this, we first studied the impact of COVID-19 spread across the world and then did a time series analysis to figure out the trend in cases and deaths across the world using FBPROPHET and ARIMA algorithms. We used the publicly available data until 21 January 2021(365 days) for building this model. The second part of this work involves predicting the case fatality rate through country specific indicators such as Socio-Economic, Health, Diet and Weather. To assess these factors, we would use Statistical Analysis and Feature Selection techniques to eliminate insignificant features and provide the most significant features that the healthcare professionals may focus on. In wake of delay in the advent of vaccines and eventual roll-out, this study aims to provide a tool that will assist the already overloaded healthcare system with its recommendations.**

Keywords -- COVID-19; Global; Pandemic; Time Series; Forecast; FBProphet; ARIMA; CFR.

I. INTRODUCTION

The year 2020 is riddled with a global pandemic named Covid-19 which has impacted every living form on this planet. Novel Coronavirus is part of the SARS family which was first identified in December 2019 in Wuhan city in the Hubei Province in China. This disease caused respiratory tract infection and, in some cases, also resulted in severe pneumonia. The disease is found to be extremely contagious by means of droplets and fomites resulting in worldwide pandemic. [1] WHO has announced it as a pandemic on March 11, 2020. [2] Since then this pandemic is spreading exponentially worldwide. As on Dec 06^{th} 2020, current active cases are 6.6million, and 1.5million deaths are reported around the world. One of the most important ways to measure the burden of COVID-19 is mortality. [3] In the current outbreak of novel COVID-19, machine learning techniques have played a vital role in finding the patterns among various countries and thus helping epidemiologists and scientists alike in their research to overcome the disease.

The forecasting of the spread of the pandemic helps to inform governments and healthcare professionals what to expect and which measures to impose, and secondly, to motivate the wider public to adhere to the measures that were imposed to decelerate the spreading lest a regrettable scenario will unfold [4]. Case Fatality Rate is a metric to evaluate the mortality associated with a disease, which can be defined as the portion of confirmed cases leading to fatality. In the modified formula by WHO, the number of closed cases is replaced by number of deaths and number of recovered cases.

$$CFR\,(\%) = (No.of\ deaths/\,No.of\ closed\ cases) * 100. \qquad (1)$$

Current data indicate that, worldwide, case fatality rate (CFR, the ratio between number of deaths and number of confirmed cases) might be around 4%. However, at the country level, CFR ranges from 0 to more than 20%. There are many possible reasons for such a variation. With this study we attempt to model the spread of the pandemic[22] and help recommend the governments and health departments on the management and availability of health infrastructure for their subjects. In absence and or delay of the vaccine this effort will help contain the disease by means of providing targeted healthcare based on individual's needs, mitigate the risks and substantiate the efforts to manage the health infrastructure.

II. LITERATURE REVIEW

Forecasting can predict the cases/deaths through the patterns associated with the time series data, this prediction provides a good understanding of the Spread of virus. This was studied using Prophet and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model forecasting model. [5] These models were able to generate forecast for confirmed cases for Canada, France, India, South Korea and UK. Even though Prophet being the procedure that is widely used for forecasting time series problems. It was not able to address the complex and varying patterns in COVID-19 Time series data, a SARIMA forecasting model is used which generates more accurate forecasts. These predictions will act as early warnings for government policy makers and help health care authorities effectively allocate resources. [5]

To take necessary actions, areas that are likely to be vulnerable should be known. Forecasting can identify the zones which have potential risk and is demonstrated in COVID-19 risk assessment in Counties of the USA [6]. Mann-Kendall and Sen's slope estimator trend analysis and homogeneity analysis (Pettitt's test) classified counties into 0-5 ranks, with 5 being 'very risky'.A Random Forest classifier was trained to classify the counties into risk zones. Socio-Economic status, household composition & disability, housing type & transportation, epidemiological factors and healthcare system factors of the counties were used for building the model and validated using Receiver Operating Characteristic (ROC)-Area Under the ROC Curve. The model achieved 90% accuracy (AUC = 0.90) during the training phase and 84% accuracy (AUC = 0.84) during the testing phase. This resulted to a Map of the USA with colors denoting the risk associated, thus enabling the officials to focus on region and factors which had led to high risk using the feature importance of the RF classifier.[6]

Forecasting[21] associated with case fatality and mortality rate can provide us valuable insights on not only with the spread of the pandemic but also handling the exposed population effectively. Forecasting deaths for different countries can capture the variable rates of fatality. To understand the irregularities pertaining to Case Fatalities for various countries, a research was conducted on Socio-Economic Factors and Health Indicators for a country's impact on the fatalities [7]. The authors have analyzed 16 potential factors which included GDP, Population, GNI, along with Health Indicators like Current Health Expenditure (CHE), Hospital beds per 10000 population, median age. Countries with at least 50 confirmed cases were selected for study at the time of research (14/03/2020). Pearson Correlation Coefficient was calculated for all factors in which 7 were observed to have very little impact on CFR. The Linear regression model built on the other 9 factors using the Ordinary Least Squares method, could capture 30.6% variability in CFR (Adjusted R-Square), GDP per capita, and the number of confirmed cases had the most impact. The best result was obtained by using the forward feature selection technique. Since only 47 countries had 50+ cases the data was insufficient, which had been the major challenge. [7]

Another interesting indicator that proved to have a significant effect on CFR was the Diet. The association between the global mortality rate of COVID-19 cases across different countries and dietary intake of different food groups was studied using an ecological study design [11]. The mortality rate is expressed in terms of CFR (Case-fatality rate). A total of ten food groups have been considered across 144 countries, data for which have been from Food and Agriculture Organization of the United Nations. The food intakes data was expressed in terms of kilocalories per person per day. The mathematical model is built using Bayesian regression model using Random-walk Metropolis-Hastings sampling. Results derived from the mathematical model's reports that COVID-19 case fatality rates were associated with food intake. The study suggests that nutritional factors available at country level could have a role in the mortality of COVID 19 pandemic.[11]

III. MATERIALS AND METHODS

The study is divided into 2 major steps, firstly understanding the spread and impact of this pandemic and then evaluating various factors such as the socio-economic features in the lethality of this global spread.

A. Predicting the global cases trend(time series)

Time Series is a series of observations taken at specified (mostly equal) time intervals. Analysis of time series helps us to predict future trends based on previously recorded values. In the Time series, we have only 2 variables, time & the variable we want to forecast, in this scenario the number of cases and number of deaths. To see how Covid-19 has impacted our world we first need to collect the past worldwide data through open-source channels such as the dataset made available by John Hopkins University. The dataset is updated on a daily basis web scraped from reliable worldwide web sources, for this research we are using the dataset until yesterday (21Jan2021). We have created 2 data frames based on countries and dates (for cases as well as deaths). We have bucketed all the cases into labels of different ranges and plot it for the entire world (As shown in below picture). Each color represents the impact of Covid-19 on each country based on cases reported.

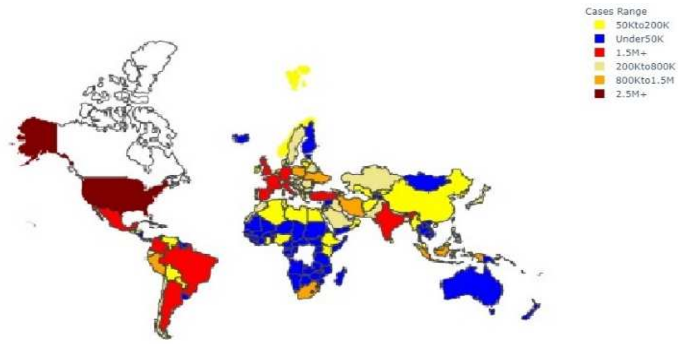

Fig 1. The map interprets the number of confirmed cases in each country (As per legend).

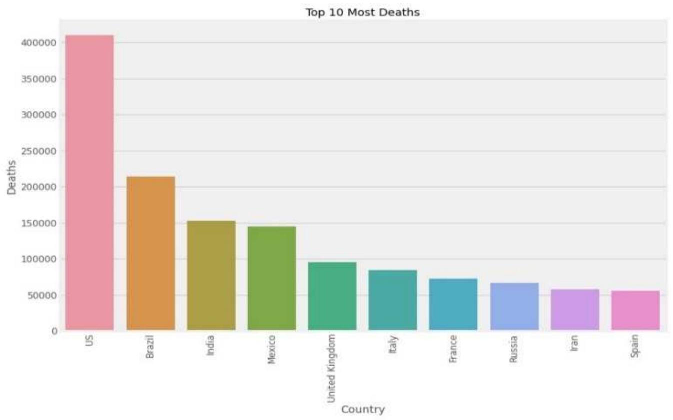

Fig 2. Histogram representation of top 10 countries based on deaths reported due to COVID_19.

For time series forecasting we have primarily used one of the 2 most popular algorithms.

1. Algorithm1- FBPROPHET:

FB Prophet is an open-source algorithm developed by Facebook for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well. Just as any time series model we first validate our data with respect to the rolling mean.

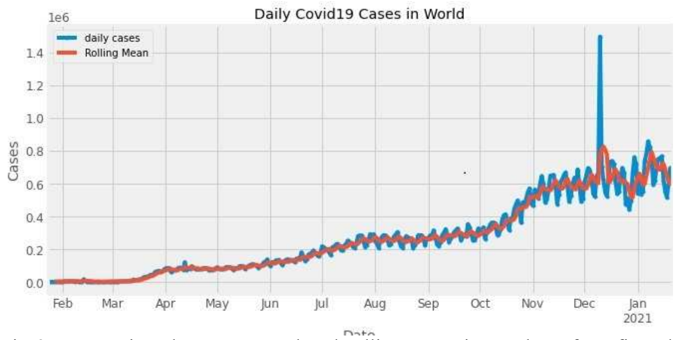

Fig 3. Comparison between actual and rolling mean in number of confirmed cases globally on daily basis.

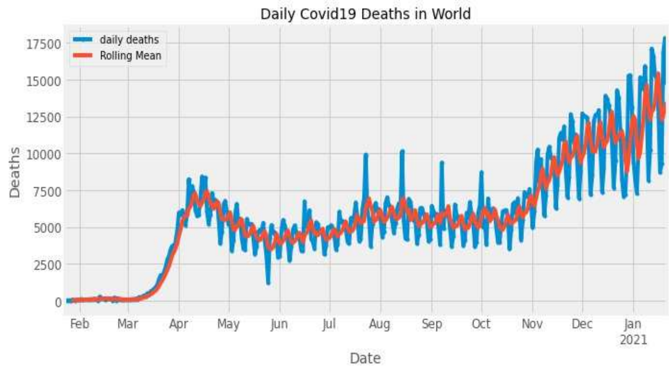

Fig 4. Comparison between actual and rolling mean in number of deaths due to covid-19 globally.

Before we look at the rolling mean one observation can be made with above graph on cases that during late December 2020 there was a sharp spike in cases observed mostly likely due to Christmas and US elections that saw mass gatherings leading to a surge in cases.

The rolling mean comparison is with the actual data that we train our model with and the moving average of that data (both confirmed cases and deaths), based on above graphs we can conclude that the rolling mean is neither overfitting nor deviating too much hence giving us the option for doing time series analysis.

2. Algorithm2 -: ARIMA:

After getting our predictions in one method we are following another algorithm named Auto Regressive Integrated Moving Average (ARIMA). ARIMA is a combination of 2 models AR (Auto Regressive) & MA (Moving Average). It has 3 hyperparameters – p (auto regressive lags), d (order of differentiation) and q (moving avg.) which respectively comes from the AR, I & MA components. The AR part is correlation between previous & current time periods. To smooth out the noise, the Moving Average part is used. The I part binds together the AR & MA parts. For the last model, ARIMA (1,1,1), a model with one AR term and one MA term is being applied to the variable

$$Z_t = X_t - X_{t-1} \qquad (2)$$

A first difference might be used to account for a linear trend in the data. The classical approach for fitting an ARIMA model is to follow the Box-Jenkins Methodology. Below is the summary of the ARIMA model developed for our modelling.

ARIMA basically involves 3 steps:

1. Model Identification: Use plots and summary statistics to identify trends, seasonality, and autoregression elements to get an idea of the amount of differencing and the size of the lag that will be required.

2. Parameter Estimation: Use a fitting procedure to find the coefficients of the regression model.

3. Model Checking: Use plots and statistical tests of the residual errors to determine the amount and type of temporal structure not captured by the model.

The process is repeated until either a desirable level of fit is achieved on the in-sample or out-of-sample observations (e.g., training or test datasets). The best fit is achieved using the lowest AIC value with respect to the best combination of p,d,q values.

To cross validate our predictions we will first split the data midway and then try to plot the predictions for the number of cases as well as deaths to see the accuracy we can achieve. Below mentioned graphs show the predicted value in red compared to the actual values in blue, showing us that our accuracy is pretty good in terms of predictions. We have split our daily data midway until June 2020.

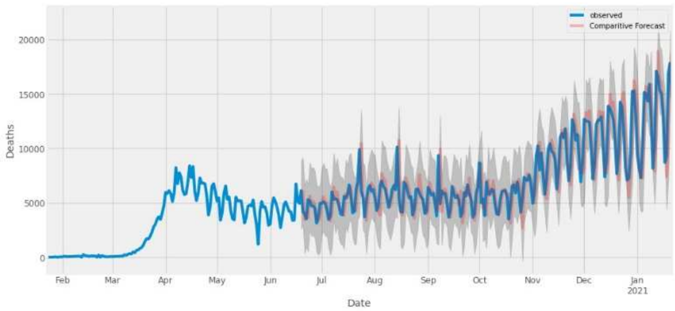

Fig 5. Cross validation between the actual data and predicted data in terms of deaths split midway over time.

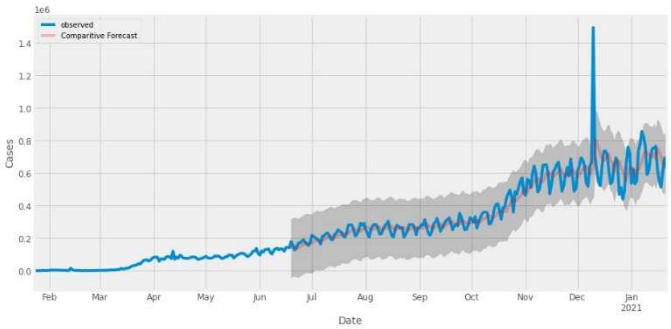

Fig 6. Cross validation between the actual data and predicted data in terms of cases split midway over time

B. Analysis of factors impacting CFR

The John Hopkins coronavirus dataset has been used to get the attributes 'confirmed cases', 'recovered cases' and 'deaths' until 15th Jan 2020 for every country. 'WHO' has recommended to calculate the CFR as ratio of deaths to sum of deaths and recovered cases, as active cases will not cause any bias in the study. Upon exploration of data, it has been observed that not all countries have followed the same standard to classify a case as recovered. For instance, in Belgium when the patient after treatment havingthe slightest of symptoms like headache were still considered as an active case. Therefore, having very low recovered cases leading to high CFRs. Hence the CFR for this study was calculated by ratio of deaths to sum of deaths and confirmed cases.

GDP, Population, Life Expectancy, GNI per capita, GDP per capita and Human capital Index (HCI) formed the socio-economic factors taken from the World Bank website. Current Health Expenditure (CHE) per capita, CHE as % of GDP, Hospital beds and number of doctors were extracted from WHO (World Health Organization). After the preprocessing and dropping countries with null values, the count came down to 136. Null values were not imputed as it will add a bias to the data.

As seen in research conducted previously in the Literature Review, CFR is treated as a regression problem. The model is expected to capture the maximum variability of the CFR for countries and in the process analyze the features which are able to explain this variability. R square is used as the metric to validate the model. Each attribute's correlation with the CFR is studied as the univariate analysis. Feature Importance was evaluated with Recursive Feature Extraction (RFE) and Forward & Backward Feature Selection methods. Linear Regression and Random Forest regressor were built on the most important features.

In the next approach, the CFR was bucketed into two classes 0 and 1, where '0' class indicates CFR less than 0.02 that is 2% of the confirmed cases lead to mortality. This approach gives the model to study the CFR by not precisely predicting the CFR. The aim is to study the features and not the precise prediction of CFR. Feature importance was evaluated the same way as the Regression, Logistic Regression and Random Forest Classifier were built to classify and CFR classes.

IV. RESULTS

The FBProphet algorithms give us an estimated prediction in terms of both cases as well as deaths on a day-to-day basis that only suggest that the number of cases as well as deaths will rise going ahead with time.

However, the weekly trend shows a sharp drop during the weekends suggesting that increased number of gatherings and decreased number of testing during the weekends that lead to more cases being reported during weekdays.

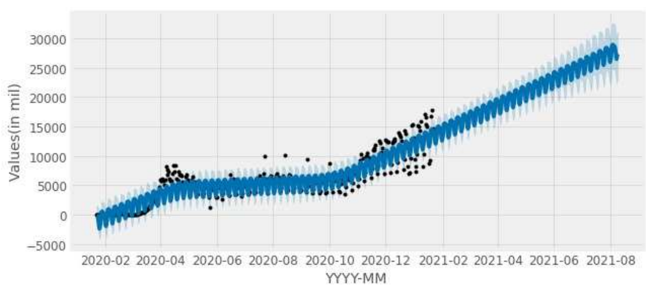

Fig 7. Getting the predictions of deaths (in thousands) for next 200 days (with black spots being the available data that was trained into model)

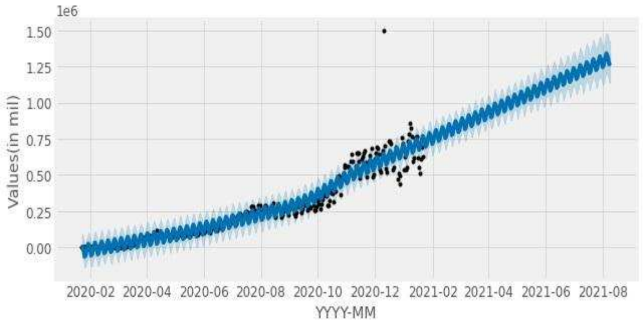

Fig 8. Getting the predictions of cases (in million) for next 200 days (with black spots being the available data that was trained into model)

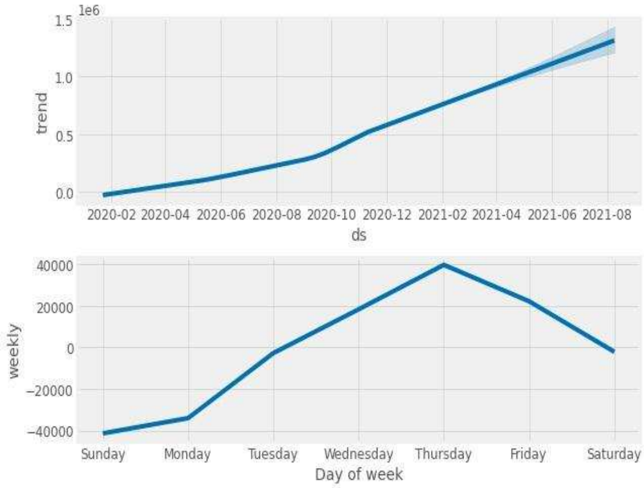

Fig 9. The monthly and weekly predictions of confirmed cases.

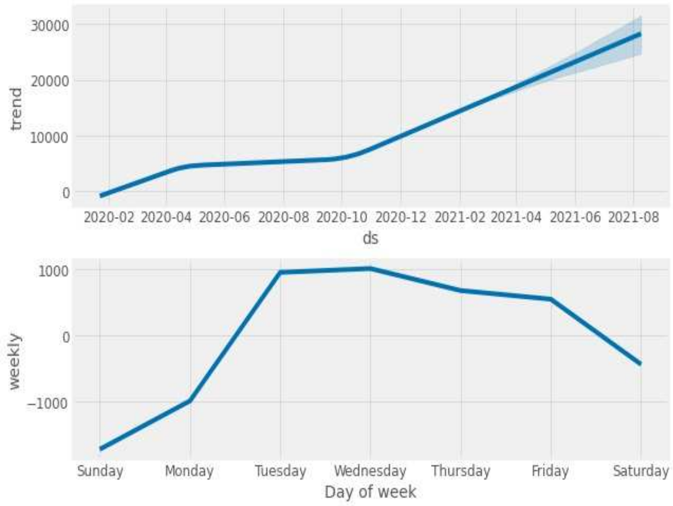

Fig 10. The monthly and weekly prediction of deaths globally.

In the case of ARIMA we can infer that even though the total number of cases in the next 200 days would increase but the number of deaths will probably see a gradual decrease mostly attributed to herd immunity

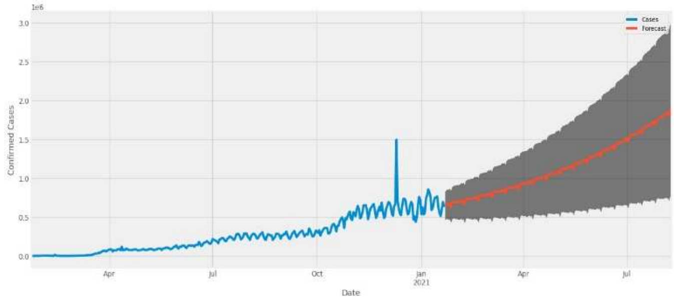

Fig 11. ARIMA prediction of total confirmed cases globally for next 200 days (Shown in red).

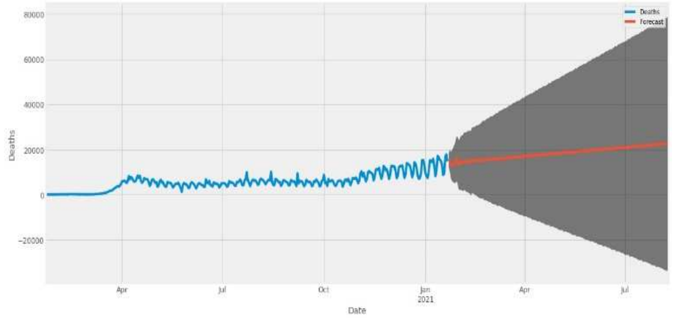

Fig 12. ARIMA prediction of total deaths globally for next 200 days (Shown in red).

The topmost impacting factors and their correlation to CFR can be seen in the below table

TABLE 1: TOP FEATURES HAVING GOOD CORRELATION WITH CFR

Attributes	Correlation	Sign
Stimulants	0.205982	-
Urban Population	0.200971	+
Miscellaneous	0.179596	-
Cereals - Excluding Beer	0.177802	+
Total Population	0.151331	+
Hospital Beds	0.146754	-
GNI/CAPITA	0.146533	-
Starchy Roots	0.140317	-
Fruits - Excluding Wine	0.116855	-
Spices	0.116002	-
HCI	0.114813	-
GDP	0.113702	+
Fish, Seafood	0.113460	-
CHE/CAPITA	0.111517	-
Milk - Excluding Butter	0.110685	-
Doctors	0.109849	-

The '+' correlation of CHE with confirmed, recovered and deaths shows that the countries with higher medical expenditure have covered more tests thus, the higher numbers. The countries with higher urban population reemphasize the above fact.

The regression analysis was conducted along with sequential feature selection with backward and forward selection. The

Linear Regression model has yielded a 0.277 R2 score which indicates the variability of the data around its mean. Random Forest Regressor was able to achieve -0.8489 R2. The table 1 and 2 represents the attributes and their influence on the Target which is the CFR. Of which It is observed that Doctors, HCI, Life Expectancy, CHE/GDP, CHE/CAPITA, Alcoholic Beverages are significant factors.

TABLE 2: FEATURES SELECTED BY BACKWARD SELECTION FOR LINEAR REGRESSION MODEL AND THEIR RESPECTIVE COEFFICIENTS.

Attributes	Coefficients	Sign
Cereals - Excluding Beer	0.012295	+
Animal Products	0.010490	+
HCI	0.008315	-
Obesity	0.008146	+
Life Expectancy	0.005329	+
DOCTORS	0.004749	-
Vegetable Oils	0.003736	+
CHE/GDP	0.003354	+
Alcoholic Beverages	0.002875	+
URBAN Population	0.002606	+

TABLE 3: FEATURES SELECTED BY BACKWARD SELECTION FOR RANDOM FOREST REGRESSOR AND THEIR FEATURE IMPORTANCE.

Attributes	Feature Importance
DOCTORS	0.140603
Spices	0.135485
Alcoholic Beverages	0.134923
CHE/GDP	0.100956
Milk - Excluding Butter	0.087006
CHE/CAPITA	0.079828
Vegetable Oils	0.078902
GNI/CAPITA	0.059623
Undernourished	0.057868
Vegetal Products	0.052789
Sugar Crops	0.040464
Animal Products	0.031553

Doctors, HCI CHE/GDP, CHE/CAPITA have Inverse effect on CFR, which implies higher values of these factors results in lower CFR and vice-versa. Life Expectancy and Alcoholic Beverages have a direct effect on CFR. The early trends of COVID19 outbreak in Italy, established that the aged population was worst affected when it comes to the infection/mortality.

To Understand the Data more and to identify underlying patterns, the classification analysis was conducted and the data was grouped by CFR <0.02.

TABLE 4: FEATURES SELECTED BY BACKWARD SELECTION FOR LOGISTIC REGRESSION AND THEIR COEFFICIENTS.

Attributes	Coefficients	Sign
DOCTORS	0.968987	-
CHE/CAPITA	0.687164	-
Starchy Roots	0.681179	-
CHE/GDP	0.552446	+
Miscellaneous	0.492156	-
Hospital Beds	0.393863	-
Meat	0.301703	+
Cereals - Excluding Beer	0.254172	-
Obesity	0.226581	+
Pulses	0.212649	-
Alcoholic Beverages	0.176067	+
Spices	0.174774	-
Undernourished	0.080433	-
GDP	0.055724	+
Animal Products	0.003265	-

The features that were selected for regression models along with hospital beds had significant effect on the classification of CFR classes. Logistic Regression was able to classify countries into low and high CFR classes with Cross validate mean accuracy of 65% and variance of 0.7%. Whereas the Random Forest Classifier had mean accuracy of 58% and variance of 3.5%.

TABLE 5: FEATURES SELECTED BY BACKWARD SELECTION FOR RANDOM FOREST CLASSIFIER AND THEIR FEATURE IMPORTANCE.

Attributes	Importance
URBAN Population	0.096616
Rural Population	0.090795
DOCTORS	0.085755
Spices	0.085121
Obesity	0.083540
Vegetal Products	0.073592
Starchy Roots	0.070095
Hospital Beds	0.069928
Cereals - Excluding Beer	0.064682
Animal fats	0.063623
Offal	0.057456
Eggs	0.055637
Vegetables	0.053458
Undernourished	0.038416
Sugar Crops	0.011286

V. DISCUSSION AND CONCLUSION

Almost a year ago, not everyone in this world was aware of the severity related to Novel Coronavirus, which has now transpired into a global pandemic and the world is still trying solutions to deal with it. This study certainly hopes to help and clarify the impact of this pandemic. There were questions related to data that has been recorded as the data only reflects based on the number of tests carried out by respective governments. This issue of low number of tests being carried out can be attributed to the lack of medical professionals and infrastructure in some countries, while other countries intentionally carried out low miscellaneous political reasons. This does to an extent hinder research in drawing out a clear picture in dealing with such pandemic.

Our research involves validating the data available to perform modelling and predictions, we have done Exploratory Data Analysis (EDA) to make sure that there are no null/NAN values in the data used, also ensuring that such values are treated appropriately(imputation), The various data types of available columns have been changed accordingly so that the predictive modelling is done without any bias or misappropriation.

With both the algorithms one thing that can certainly be predicted that there will be a gradual increase in the total number of cases globally even though the number of deaths won't see a huge spike. During the regression and classification study, it was found that there are common diet and socio-economic features that affect CFR. Countries that are predicted to have higher CFR could contain the increasing mortality by investing in healthcare infrastructure. The use of certain ingredients in the diet such as spices can contribute to reducing mortality. The diet factors can explain the low CFR for India and countries that use such ingredients in their cuisines. The timing of this study has not taken the exception of Covid-19 vaccine now available from 5 vaccine makers and vaccination drives set to begin across 150 countries, we can expect to see a sharp drop in the number of cases and deaths reported given the effectiveness of the vaccine. However, there are various mutations of this virus that have developed into regional strands showing divergent characteristics moving ahead, we have to wait and see how mankind would learn and respond to the pandemic that has already affected lives in a big way.

VI. References

[1] Estimating mortality from COVID-19. (n.d.). Retrieved February 1, 2021, from Who.int website: https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19

[2] Sarkar, K., Khajanchi, S., & Nieto, J. J. (2020). Modeling and forecasting the COVID-19 pandemic in India. *Chaos, Solitons, and Fractals*, *139*(110049), 110049.

[3] Coronavirus update (live): 103,557,049 cases and 2,238,513 deaths from COVID-19 virus pandemic - worldometer. (n.d.). Retrieved February 1, 2021, from Worldometers.info website: https://www.worldometers.info/coronavirus/

[4] Perc, M., Gorišek Miksić, N., Slavinec, M., & Stožer, A. (2020). Forecasting COVID-19. *Frontiers in Physics*, *8*. doi:10.3389/fphy.2020.00127

[5] Chakraborty, T., & Ghosh, I. (2020). Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. *Chaos, Solitons, and Fractals*, *135*(109850), 109850.

[6] (N.d.). Retrieved February 1, 2021, from Researchgate.net website: https://www.researchgate.net/publication/344360045_Using_Machine_Learning_to_Develop_a_Novel_COVID-19_Vulnerability_Index_C19VI

[7] Asfahan, S., Shahul, A., Chawla, G., Dutt, N., Niwas, R., & Gupta, N. (2020). Early trends of socio-economic and health indicators influencing case fatality rate of COVID-19 pandemic. *Monaldi Archives for Chest Disease*, *90*(3), 451–457.

[8] Yadaw, A. S., Li, Y.-C., Bose, S., Iyengar, R., Bunyavanich, S., & Pandey, G. (2020). Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *The Lancet. Digital Health*, *2*(10), e516–e525.

[9] Khafaie, M. A., & Rahim, F. (2020). Cross-country comparison of case fatality rates of COVID-19/SARS-COV-2. *Osong Public Health and Research Perspectives*, *11*(2), 74–80.

[10] Malki, Z., Atlam, E.-S., Hassanien, A. E., Dagnew, G., Elhosseini, M. A., & Gad, I. (2020). Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. *Chaos, Solitons, and Fractals*, *138*(110137), 110137.

[11] Eltoukhy, A. E. E., Shaban, I. A., Chan, F. T. S., & Abdel-Aal, M. A. M. (2020). Data analytics for predicting COVID-19 cases in top affected countries: Observations and recommendations. *International Journal of Environmental Research and Public Health*, *17*(19), 7080.

[12] Brownlee, J. (2017, January 8). How to create an ARIMA model for time series forecasting in python. Retrieved February 1, 2021, from Machinelearningmastery.com website: https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/.

[13] Tandon, H., Ranjan, P., Chakraborty, T., & Suhag, V. (2020). Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future. Retrieved from http://arxiv.org/abs/2004.07859

[14] fbprophet. (n.d.). Retrieved February 1, 2021, from Pypi.org website: https://pypi.org/project/fbprophet/

[15] freespirit. (2018, August 24). Time series for beginners with ARIMA. Retrieved February 1, 2021, from Kaggle.com website: https://www.kaggle.com/freespirit08/time-series-for-beginners-with-arima

[16] Lesson 3: Identifying and Estimating ARIMA models; Using ARIMA models to forecast future values. (n.d.). Retrieved February 1, 2021, from Psu.edu website: https://online.stat.psu.edu/stat510/book/export/html/665.

[17] Where, Z. X. (n.d.). A generalization of ARMA models which incorporates a wide class of nonstation- ary TS is obtained by introducing the differencing into the model. The simplest example of a nonstationary process which reduces to a stationary one after dif- ferencing is Random Walk. As we have seen in Section 4.5.2 Random Walk is a nonstationary AR(1) process with the value of the parameter φ equal to 1, that is the model is given by. Retrieved February 1, 2021, from Qmul.ac.uk website: http://www.maths.qmul.ac.uk/~bb/TimeSeries/TS_Chapter7.pdf

[18] COVID19-India API. (n.d.). Retrieved February 1, 2021, from Covid19india.org website: https://api.covid19india.org/documentation/csv/

[19] Prophet. (n.d.). Retrieved February 1, 2021, from Github.io website: https://facebook.github.io/prophet/

[20] Home - johns Hopkins Coronavirus resource center. (n.d.). Retrieved February 1, 2021, from Jhu.edu website: https://coronavirus.jhu.edu/

[21] N. Darapaneni, D. Reddy, A. R. Paduri, P. Acharya, and H. S. Nithin, "Forecasting of COVID-19 in India using ARIMA model," in 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2020, pp. 0894–0899.

[22] N. Darapaneni et al., "COVID-19 Infection Dynamics for India-Forecasting the Disease using SIR models," in 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), 2020, pp. 387–392.