COVID-19: VISUALIZATIONS AND FORECASTING

Aastha Gupta[1], Mahima Verma[2], Sakshi Bisht[3], Smriti Singh[4], Gunjan Chugh[5]

[1,2,3,4]*Student, Dr. Akhilesh Das Gupta Institute of Technology and Management*
[5]*Assistant Professor, Department of Information Technology, Dr. Akhilesh Das Gupta Institute of Technology and Management*

---***---

Abstract - *Coronavirus disease (COVID-19) has made a huge impact on the world and has infected more than 49.9 million people around the globe. This is an inflammatory disease highly affecting the human community which has also opened new challenges for the research community. However, predicting the trend of COVID-19 is an important challenge. Here in this study, we have focused on the investigation, visualization and prediction about the COVID-2019. With the help of numerous models in machine learning, it would converge on predicting the multiple outcomes of the coronavirus and comparing its global threat of annihilation. It forecasts the number of confirmed cases, deaths, and recovered cases in India using the prophet model and also visualizes the time-series data set on COVID-19 cases. Furthermore, this paper includes the case study of a few countries including India. This study helps in anticipating and estimating the not so distant future.*

Key Words: Coronavirus, Machine Learning, Prediction, Visualizations, Prophet

1. INTRODUCTION

The year 2020 will be forever marked in history by the worldwide outbreak of the pandemic caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2).[1] The novel Coronavirus designated SARS-CoV-2 appeared in December 2019 to initiate a plague of respiratory illness mentioned as COVID-19 which proved itself as a difficult illness which can make appearance in various forms and levels of severity starting from mild to severe with the danger of organ failure and death. From mild, self-restraint tract illness to severe progressive pneumonia, multiorgan failure, and death.[2] The virus began to spread globally, the World Health Organization (WHO) declared on the 30th of January 2020 that COVID-19 was a Public Health Emergency of International Concern (PHEIC). The virus outbreak has changed the life-style of everyone around the world, forcing governments to mandate lockdowns, recommend self-isolation, stipulate work-from-home policies, instigate strict social distancing criteria, and deploy emergency health responses - the latter including new infrastructure for the treatment and mass testing of the population at large. All these measures are directed towards decreasing the speed of spread of the virus and resulting in a so-called 'flattening of the curve' until an approved vaccine/treatment is developed.

COVID-19 is more infectious compared to other known viruses (such as SARS and MERS) albeit with a lower death rate. Furthermore, a COVID-19 carrier is often contagious without experiencing any symptoms. Thus, by the time the carrier actually tests positive, they may have already spread the virus to many others who came in contact with them.[1]

As per the World Health Organization, no vaccine and antiviral treatments are yet available for this virus, and medical organizations are trying hard to find out the vaccine for this novel coronavirus. However, even after fast-tracking the standard vaccine period of 5–10 years, the vaccine may take a minimum of 18–24 months before it's available and should take further longer to produce it enough for the bulk of the planet. Also, we do not know how long a vaccine would stay effective because the virus mutates. Every effort is formed to hamper the spread of the coronavirus and prepare medical response systems to tackle the rise in patient loads and to guard the front-line medical staff with adequate supplies of personal protective gear like personal protective equipment (PPE), masks, and other essentials.

In this paper, we have visualized and reported all continent and country wise confirmed cases, recovered cases, deaths and active cases. Further Top 10 countries having maximum number of confirmed, active, death and recovered cases are visualized. Then correlation analysis and visualizations on the world map are carried out, finally leading to various visualizations and predictions of Confirmed, Death and Recovered cases of India.

2. LITERATURE REVIEW

Determining appropriate approaches to succeed in finding solutions for the COVID-19 related problems has received a great deal of attention. However, another huge problem that researchers and decision-makers have to deal with is the ever-increasing volume of data, known as big data, that challenges them in the process of fighting against the virus.

As per John Hopkins University, 4,563,458 confirmed cases of COVID-19 were reported as of 16-May-20. India contributes 1.9% with 86,508 cases and features a death rate of 3.2% with 0.2 deaths per 100k population.[3] All countries are attempting to save lots of their people's lives by implementing measures like travel restrictions, quarantines, event postponements and cancellations,

social distancing, testing, hard and soft lockdowns. More than the lives this virus has taken, the economic and social impact is far more disastrous and especially for developing and underdeveloped countries.[3]

Machine learning (ML) has proved itself as a bodacious field of study over the last decade by solving many very complex and sophisticated real-world problems. One of the most significant areas of ML is forecasting, plentiful standard ML algorithms are being used in this area to guide the future course of actions needed in many application areas including weather forecasting, disease forecasting, stock market forecasting as well as a disease diagnosis.

According to the transmission characteristics of an epidemic at different stages, this paper focuses on predicting the future COVID-19 cases uses prophet procedure model on the COVID-19 provided data by different sources to anticipate the epidemiological example of the ailment and pace of COVID-2019 cases in India and other parts of the world that is being affected by a coronavirus.

3. PROPOSED WORK

3.1 Data set

This study provides an approach to analyzing the current situation of covid-19 around the globe, the data helps in evaluating which country is making progress to its goal to prevent coronavirus. We explore datasets with different characteristics like recovered, death, and confirmed cases (Fig-1). The data on confirmed cases only becomes meaningful when it can be interpreted as how much a country is testing the cases.

Here we statistically analyzed our dataset with various methods of data analysis and visualized those data to provide a proper understanding of the COVID-19 outbreak worldwide. Our exploit analysis is carried out on Kaggle covid-19 dataset and World Health Organization (WHO) with the 2019-20 coronavirus dataset.[4] Here, we present an effort to visualize and analyze the results. COVID-19 has so far propagated nearly 177 Countries/Regions and 32 Territories have been registered. Using time-series data, which estimates the number of individual cases, such as confirmed infected, deaths, and recovered around the globe.[6]

Reporting Country/ Territory/Area*	Total confirmed * cases	Total confirmed new cases	Total deaths	Total new deaths	Transmission classification*	Days since last reported case
Western Pacific Region						
China	82930	55	3338	3	Local transmission	0
Republic of Korea	10237	81	183	6	Local transmission	0
Australia	5635	181	34	6	Local transmission	0
Malaysia	3483	150	57	4	Local transmission	0
Japan	3271	351	70	1	Local transmission	0
Philippines	3094	76	144	8	Local transmission	0
Singapore	1189	75	5	0	Local transmission	0
New Zealand	872	48	1	0	Local transmission	0
Viet Nam	240	1	0	0	Local transmission	0
Brunei Darussalam	135	1	1	0	Local transmission	0
Cambodia	114	0	0	0	Local transmission	1
Mongolia	14	0	0	0	Imported cases only	3
Fiji	12	5	0	0	Local transmission	0
Lao People's Democratic Republic	10	0	0	0	Local transmission	3
Papua New Guinea	1	0	0	0	Imported cases only	15
Territories**						
Guam	93	9	4	1	Local transmission	0
French Polynesia	40	1	0	0	Local transmission	0
New Caledonia	18	0	0	0	Local transmission	2
Northern Mariana Islands (Commonwealth of the)	8	0	1	0	Local transmission	2
European Region						
Spain	124736	7026	11744	809	Local transmission	0
Italy	124632	4805	15362	681	Local transmission	0
Germany	91714	5936	1342	184	Local transmission	0
France	67757	4221	7546	1053	Local transmission	0
The United Kingdom	41907	3735	4313	708	Local transmission	0
Turkey	23934	3013	501	76	Local transmission	0
Switzerland	20489	783	666	59	Local transmission	0
Belgium	18431	1661	1283	140	Local transmission	0
Netherlands	16627	904	1651	164	Local transmission	0
Austria	11766	241	186	18	Local transmission	0
Portugal	10524	638	266	20	Local transmission	0
Israel	7589	559	42	6	Local transmission	0
Sweden	6443	365	373	40	Local transmission	0
Norway	5510	302	50	6	Local transmission	0
Russian Federation	4731	582	43	9	Local transmission	0
Ireland	4604	331	137	17	Local transmission	0

Fig -1: Data set used for visualization of covid-19 cases around the world[4]

Name	Cases - cumulative total	Cases - newly reported in last 24 hours	Deaths - cumulative total	Deaths - newly reported in last 24 hours	Transmission Classification
Global	48,551,059	591,869	1,231,247	9,073	
United States of A…	9,387,978	106,050	232,166	1,132	Community transmission
India	8,411,724	47,638	124,985	670	Clusters of cases
Brazil	5,590,025	23,976	161,106	610	Community transmission
Russian Federation	1,733,440	20,582	29,887	378	Clusters of cases
France	1,567,332	58,017	38,718	363	Community transmission
Spain	1,306,316	21,908	38,486	368	Community transmission
Argentina	1,205,928	9,516	32,520	468	Community transmission

Fig -2: These are the top seven countries affected due to coronavirus till 6 November 2020[4]

The data set (Fig-1) represents the early confirmed cases and the death rate around the world but this was the early stage effect of the virus and with time the number of death and confirmed cases increased, in the next figure (Fig-2), we can clearly see the sudden surges of cases in the country such as United States of America, India, and brazil. From 30 December 2019 to 11 October 2020, over 37 million COVID-19 cases and 1 million deaths have been reported globally. Nearly half of these cases (48%) and deaths (55%) [4]continue to be outlined in the Region of the America with the United States of America, Brazil, and Argentina accounting for the maximum numbers of new cases and deaths in the region.

The research also addresses analyzing the current situation of covid-19 in the different states of India. In India, 100,340 number of confirmed cases and 3155 death cases due to COVID-19 were reported on May 18, 2020.[4] Due to the absence of a specific vaccine, non-pharmacological interventions including social distancing, contact tracing are essential for ending the worldwide COVID-19.[6]

The reason cited for the sudden spurt in cases is due to the reopening of manufacturing and economic activities across the country, people have started moving out of their houses, even as usual crowds are once again returned at the market places. There are other reasons as well that are the carelessness among the people in adopting precautions like wearing face-masks, maintaining social distancing, and regular use of hand-sanitizers.

3.2 Algorithm

This study is focused on visualizations and live forecasting of COVID-19 confirmed cases, death cases, and recovered cases for the next two months that is 60 days from 17th November 2020 to 16th January 2021 using Prophet model.

Prophet is an open-source tool from Facebook used for forecasting time series data which helps businesses understand and possibly predict the market,[4] whereas prophet used here is to implement a procedure for forecasting time series data supported 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 a number of other historical data. Prophet is robust to missing data and shifts within the trend, and handles outliners typically well.[5]

The general idea of the model is analogous to a generalized additive model. The "Prophet Equation" fits, trends, seasonality and holidays. This is given by,

$$y(t) = g(t) + s(t) + h(t) + e(t) \text{ [5]}$$

where g(t) refers to changes over a long period of time or trend, s(t) refers to seasonality that is periodic or short-term changes, h(t) refers to effects of holidays to the forecast, e(t) refers to the unconditional changes that's specific to a business or an individual or a circumstance. It is also called the error term and y(t) is the forecast.[5-6]

Prophet provides us with two models(though the newer models can be written or extended according to requirements). One is the logistic growth model while the other one is a piecewise linear model. By default, the Prophet uses a piecewise linear model, but it can be changed by specifying the model according to the requirement. Choosing a model is delicate as it is dependent on a variety of factors such as Date, time, State/Union Territory, Confirmed Foreign National, Deaths cases, cured/recovered cases and Confirmed cases and growth rate, etc. If the data to be forecasted has saturating and non-linear data(grows non-linearly and after reaching the saturation shows little to no growth or shrink and only exhibits some seasonal variations), then logistic growth model is the best option.[5-6]

The model was trained on the first 210 days on the time series data set of the India's current confirmed and death cases due to the pandemic starting from 30th January 2020 and then tested on further 60 days till 1st November 2020, in order to predict further number of cases from 17th November 2020 to 16th January 2021. This prediction system can be very helpful in decision making to handle the present scenario to guide early interventions to manage these diseases very effectively.[9]

3.3 CASE STUDY

This research examines the current and future situation of the world due to coronavirus. Here the study not only revolves around the analysis and visualizations of the World and India, but also the prediction of the confirmed, death and recovered cases of India.

This paper visualizes the global impact of covid-19. Fig-3 describes the plot of top 10 countries having maximum number of confirmed, death, active and recovered COVID-19 cases. Dealing with the unexpected challenges caused by COVID-19 pandemic has taken a significant toll on people all around the world. At the time of writing this Special Feature, there are over 35.9M confirmed cases of COVID-19 across the globe.[6]

The pandemic rages unabated across continents, North Africa turned out to be most affected by the novel coronavirus disease (COVID-19) pandemic across the continent, with more than half of Africa's reported cases and deaths are recorded from this region. Approximately 299 deaths — or 68 percent of all deaths in Africa — were reported in six countries of the region, with the number of cases rising sharply to 4,204.[7]

This study uses various python libraries like pandas for COVID-19 data analysis and manipulation, matplotlib for simple and powerful 2-D plots of various confirmed, recovered and deceased cases, folium for visualization on Leaflet World and India map, calmap for calendar heatmaps, pycountry_convert for providing conversion functions between ISO country names, country-codes, and continent names and plotly for controlling the appearance of the base map or generating the heat map for global COVID-19 cases.
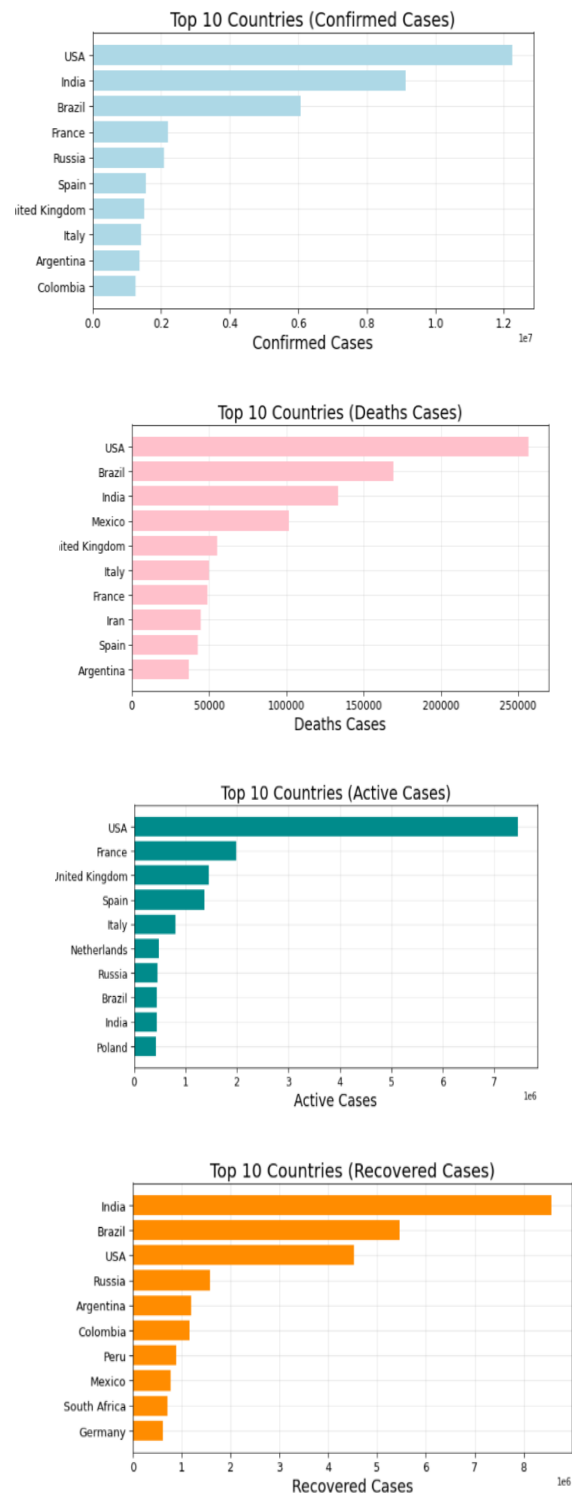

Fig -3: Plot showing confirmed, death, active and recovered cases of top 10 countries

In India, the global pandemic COVID-19 first made its presence on 30 January 2020. The country has taken several measures to contain the spread of pandemic which includes lock-down of the entire country for nearly forty

days.[8] and relaxed in three phases since June but this showed very minimal effects on the covid-19 cases in India.

Fig-4 describes the total number of tests done in India till 20th November 2020. Fig-5 and Fig-6 describe the 10 most affected states in India due to COVID-19. Fig-7 describes the total number of confirmed, recovered, death and active cases till 16th November 2020 using matplotlib library and functions like bar, barh, background_gradient etc.
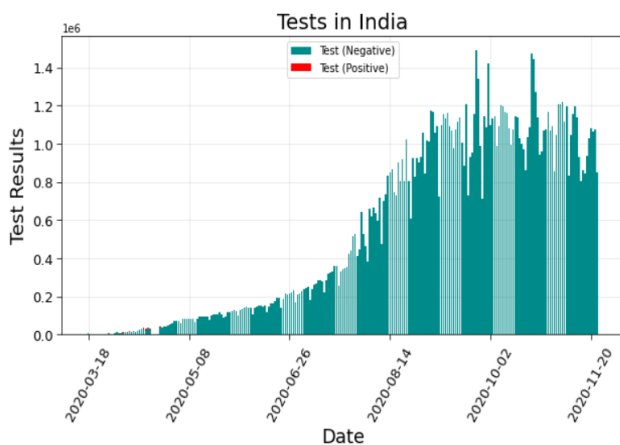

Fig -4: Plot showing total tests done till date (20th November 2020) in India
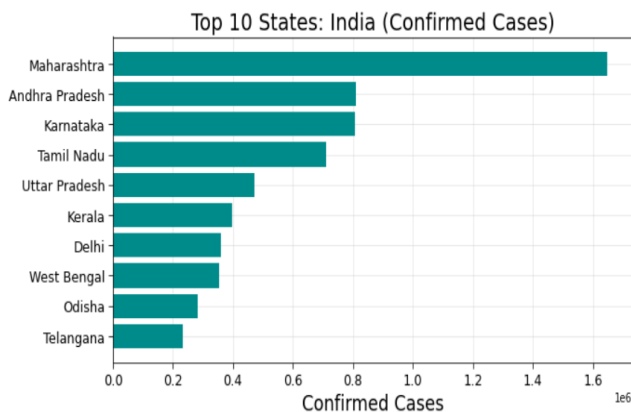
Fig -5: Top 10 states having maximum number of confirmed cases in India
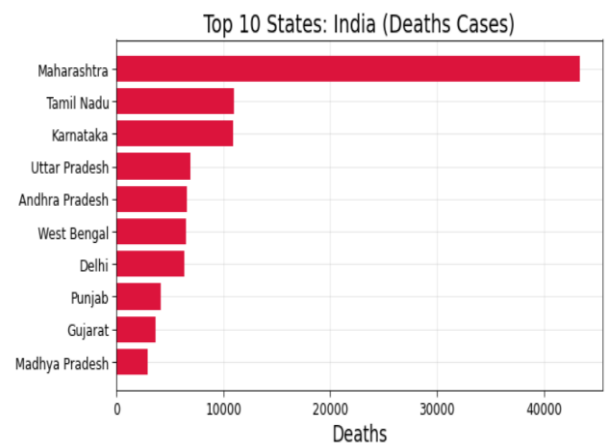
Fig -6: Top 10 states having maximum number of Death cases in India

	confirmed	recovered	deaths	active	Mortality Rate (per 100)
Total	7945975.000000	7198877.000000	119538.000000	626192.000000	1.500000

Fig -7: The total number of confirmed, death, recovered, and active cases in India till 16th November 2020[4]

4. RESULTS

Fig-8 and Fig-9 are based on visualizing the number of confirmed cases, death cases, recovered cases and their incidence & mortality rate around the world which is the first module of this research. Fig-10 visualizes the cases in India.

	Confirmed	Deaths	Recovered	Active	Incident_Rate	Mortality Rate (per 100)
0	59225734	1397312	37954816	19002612	208874.32	2.36

Fig -8: Study visualizing the total number of confirmed, death, recovered, and active cases across the world
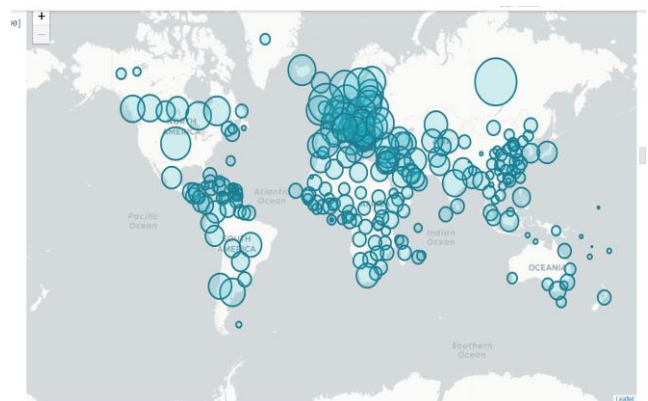

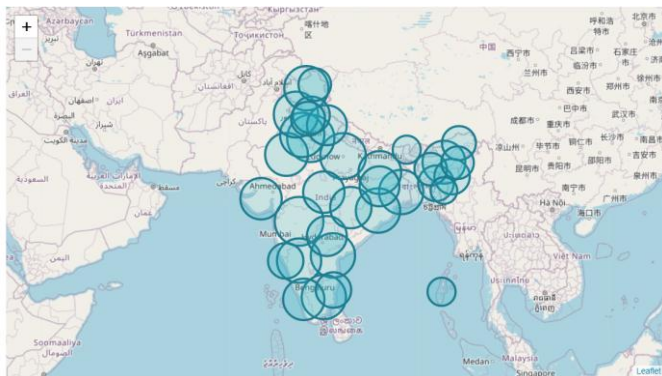**Fig-9:** Visualization of Covid-19 cases on World map

Fig-10: Visualization of Covid-19 cases on India map

Fig-11 to Fig-16 is forecasting the number of confirmed, death and recovered cases in India using Prophet Model for 60 days starting from 17th November 2020 to 16th January 2021 resulted in the average precision of 94% with the average MAPE (Mean Absolute Percent Error) value of 6 which is the second module of this research paper.
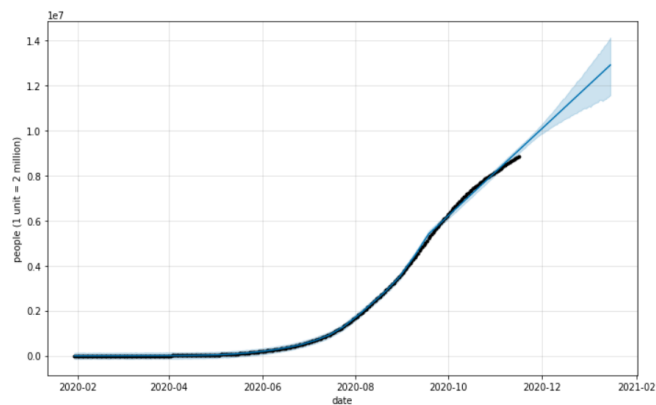

Fig-11: Prediction of number of confirmed cases in India with Mean Absolute Percent Error of 4.056 and precision of 95.94 %
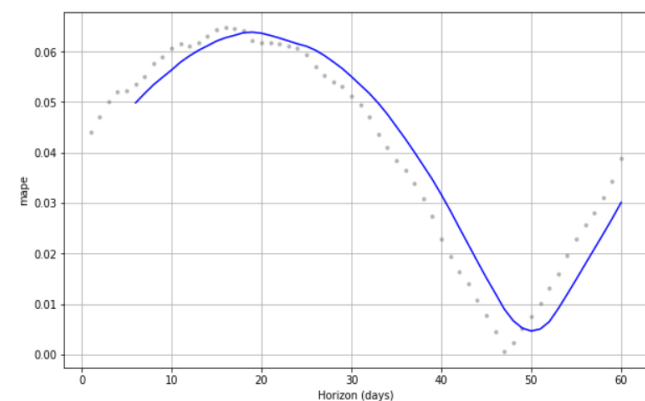
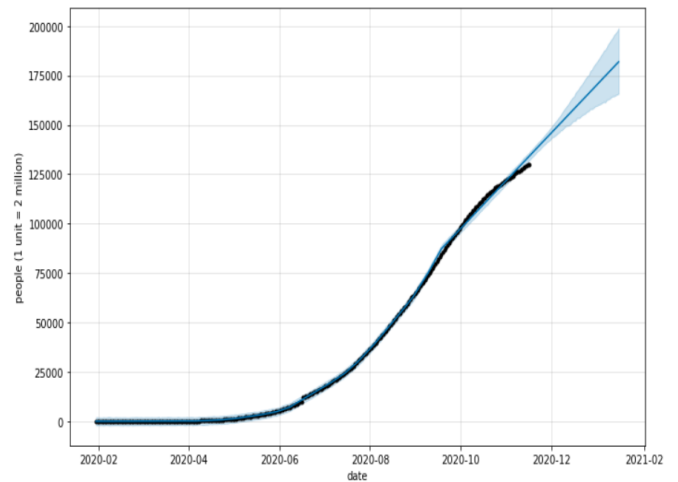**Fig-12:** Plot of cross validation metric of predicted confirmed cases in India due to Covid-19

Fig-13: Prediction of number of death cases in India with Mean Absolute Percent Error of 3.130 and precision of 96.87 %
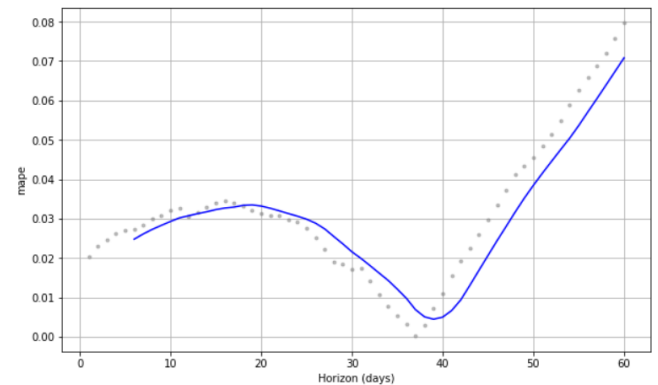

Fig-14: Plot of cross validation metric of predicted death cases in India due to Covid-19
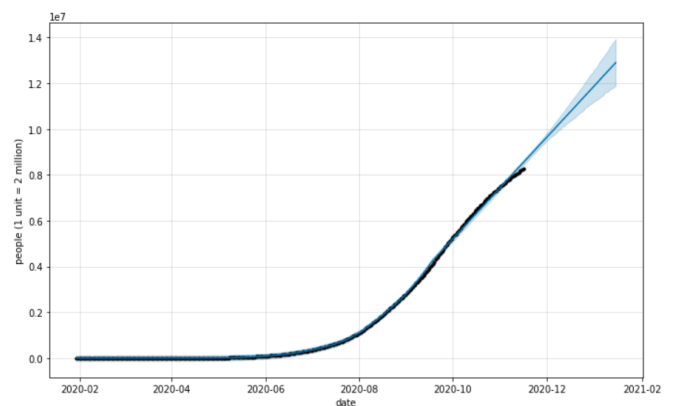

Fig-15: Prediction of number of recovered cases in India with Mean Absolute Percent Error of 10 and precision of 90 %
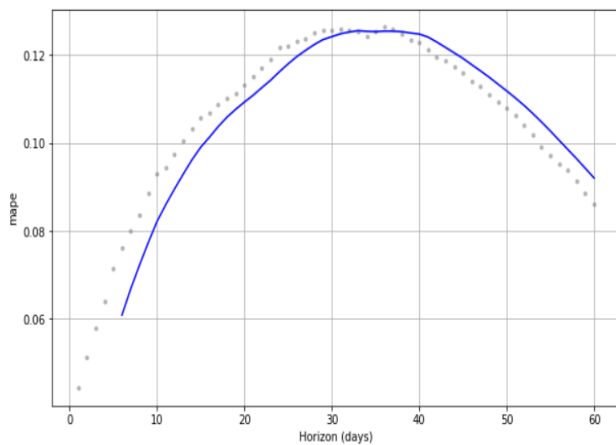
Fig-16: Plot of cross validation metric of predicted recovered cases in India due to Covid-19

5. CONCLUSIONS

In this paper, we have predicted the number of confirmed, death and recovered COVID-19 cases in India using the Prophet model. The precision of our model is 94%. Further, it includes a user-friendly data visualization model of India as well as the world. The dataset utilized in this research work of COVID-19 is from Kaggle and World Health Organization (WHO) that's a specialized agency of the United Nations liable for international public health. We try to predict the future cases which may increase around the near future and analyze the COVID-19 cases on the basis of a decrease or increases in the recovery cases in India. As a caveat, this is a data analysis and visualization approach of COVID-19 that is rapidly evolving. However, knowledge about this novel SARS-CoV-2 virus remains limited among general people around the globe. Raw data released by various sources are not adequately capable to provide an informative understanding of COVID-19, caused by SARS-CoV-2. This research work will be more effective to understand the epidemic outbreak of this severe disease. Visualization models like map view and table view discussed in this paper provide an interactive interface and visualize each and every raw fact in a comprehensive manner.

ACKNOWLEDGEMENT

REFERENCES

[1] AHMED N, MICHELIN RA, XUE W, RUJ S, MALANEY R, KANHERE SS, SENEVIRATNE A, HU W, JANICKE H, JHA SK. A SURVEY OF COVID-19 CONTACT TRACING APPS. IEEE ACCESS. 2020 JUL 20;8:134577-601.

[2] JAMSHIDI M, LALBAKHSH A, TALLA J, PEROUTKA Z, HADJILOOEI F, LALBAKHSH P, JAMSHIDI M, LA SPADA L, MIRMOZAFARI M, DEHGHANI M, SABET A. ARTIFICIAL INTELLIGENCE AND COVID-19: DEEP LEARNING APPROACHES FOR DIAGNOSIS AND TREATMENT. IEEE ACCESS. 2020 JUN 12;8:109581-95.

[3] ARORA P, KUMAR H, PANIGRAHI BK. PREDICTION AND ANALYSIS OF COVID-19 POSITIVE CASES USING DEEP LEARNING MODELS: A DESCRIPTIVE CASE STUDY OF INDIA. CHAOS, SOLITONS & FRACTALS. 2020 OCT 1;139:110017.

[4] CORONAVIRUS DISEASE 2019 (COVID-19) HTTPS://COVID19.WHO.INT/?GCLID=CJWKCAIA-F78BRBBEIWATKRRBMO06GGOLR2J0HXTKGISGPWGDQVOCROFL4OTWMCH7815XBJICZHW2BOCRRKQAVD_BWE

[5] FORECASTING AT SCALE HTTPS://FACEBOOK.GITHUB.IO/PROPHET/

[6] TIME SERIES ANALYSIS USING FACEBOOK PROPHET HTTPS://WWW.GEEKSFORGEEKS.ORG/TIME-SERIES-ANALYSIS-USING-FACEBOOK-PROPHET/

[7] COVID-19: NORTH AFRICA MOST AFFECTED IN ENTIRE CONTINENT WITH 68% OF ALL DEATHS HTTPS://WWW.DOWNTOEARTH.ORG.IN/NEWS

[8] NAGARAJA SB, KELAMANE S, KUNJATHUR SM, GURUM S. UPROOTED BY COVID PANDEMIC: NATIONAL TB ELIMINATION PROGRAMME NEEDS ACCELERATION! THE INDIAN JOURNAL OF TUBERCULOSIS. 2020 AUG 28.

[9] RUSTAM F, RESHI AA, MEHMOOD A, ULLAH S, ON B, ASLAM W, CHOI GS. COVID-19 FUTURE FORECASTING USING SUPERVISED MACHINE LEARNING MODELS. IEEE ACCESS. 2020 MAY 25.