Forecasting the Next Pandemic: Covid19 using Machine Learning

Vani Malagar, Akshita Sharma

Abstract:

COVID-19 has been declared as a pandemic in 2020 and smart technologies have demonstrated their noteworthiness to forcast the consequences of Covid-19 outbreak and to improve the ways of handling such kind of pandemic in future. Smart technologies and machine learning (ML) models have been utilized in numerous applications to predict the impact of Covid-19 on different types of patients with respect to age, health, gender, and medical history. The study has considered various parameters to draw inferences from the data collected worldwide. This experimental study reveals the ability of ML models to figure out the number of upcoming patients infected by COVID-19. The standard dataset from KAGGLE has been used for the research study for predicting the Covid-19 infected patients around the world. In this paper, Machine learning techniques are applied to build the model forpredicting the positive cases from the worst-hit countries with the latest available datasets. The research paper provides insights into theimpact of COVID-19 worldwide. The proposed methods also assist in predicting the futuristic infected cases of Covid-19.

Keywords: Covid19, Linear Regression,LASSO, Exponential smoothening ,Forecasting.

I. INTRODUCTION

As indicated in [1], affirmed corona-virus infection 2019 (COVID-19) cases are developing exponentially in many nations around the globe. In US and Brazil, the pandemic is overburdening the social framework [2], and as the current situation won't be controlled then other nationas will face the same consequences as faced by European nations and the United States. COVID-19 has impacted our lives tremendously in all respects [3,4,5,6,7]. It is difficultfor the governments what measures should been forced to control the pandemic, and furthermore, to take the more extensive open to hold fast measures that were forced to decelerate the spreading in case an unfortunate situation will unfurl [8, 9]. Exploration on datasets has a long and productive history in measurable data science and the epidemiology [10, 11]. Basic numerical models that depict the pith of pandemic spreading can be utilized to fit the information with an oversee that count the boundaries, and the acquired data now then be utilized to fullfill educated expectations. As of now,the investigation has collected overpowering proof for unpredictable and heterogeneous data designs in interpersonal organizations [12–16]. The assumption now is a key job in deciding the conduct of balance and non-balance frameworks as a rule, and the spreading of pandemics and finding ideal techniques specifically in that domain.

Interdisciplinary investigations at the interface of measurable material science, data science, and the study of disease transmission, driven by measures of information updation in our well-being and lifestyle, have offered ascend to advance the study of disease transmission [17] and to look into the hypothesis on complex systems [10]. From traditional models that accept very much blended populaces, to the later models that represent the utilization of the input and the structure of our interpersonal organizations, we have progressed significantly in better understanding ailment transmission and infection elements. We are currently ready to utilize this information to create powerful anticipation systems [11], and all the more comprehensively, we can utilize the cooperative energies between these various fields of exploration to improve our lives and social orders [18, 19].

In any case, in the midst of criticalness even the simplest model can be excessively muddled, and the areas between various fields of exploration can appear resolve the openings. In this paper, we hence present a basic iterative strategy to figure the increasing cases of COVID-19, under the suspicion that administrative information is genuine and honest. The objective isn't to make progress toward careful precision nor to introduce our technique as the best in class, however just to give first experiences and

rules on rudimentary standards. The work further motivate to yield increasingly detailed and precise forecast strategies.

In December 2019, the Chinese authorities notified the virus was spreading world in the following months, it spreads to other countries with cases doubling within days. This virus is the severe acute respiratory syndrome related to corona-virus 2 and everyone simply called as covid-19. A Virus was spreading and actually happens when it infects a human. A virus is just a hole around genetic material and a few spike proteins, RNA and Liquid Envelope by entering the living but it's still on setting i.e. how long has infection last when people call or if you touch someone. This is almost when you'll face towards the infected ones, rubbing your eyes or nose that virus sit at the places and enters through the sensitive positions and reaches to its destinations i.e. lungs where it can have the most dramatic effects.

AI (ML) has substantiated itself as a conspicuous field of study in the course ofthe most recent decade byunravelling numerous exceptionally sophisticated real issues. The application included practically all this present technical areas, for example, human services, self-governing vehicle (SGV), Natural Processing Language (NLP),business applications,Automatic robots, gaming, Weather Prediction, voice and so on. Machine learning is regularly founded on experimentation technique that includes customary calculations, which adheres to the programming directions dependent on parameters [1]. One of the most technical areas of ML is estimating [2], various standard ML algorithms that have been utilized around the different applications in order to manage the future course of activities required in numerous application including climate determining, illness forecasting, financial exchange forecasting and so on.Different neural system models have wide relevance in foreseeing the status of patients later on with a particular infection [3]. There are various studies performed for the expectation of various problems where AI strategies are being utilised,for example, coronary artery infection [4], cardiovascular illness forecast [5], and breast malignancy forecast [6]. Specifically, the examination [7] is centred on live anticipating of COVID-19 affirmed cases and study [8] is likewise centred to estimate the COVID-19 flare-up and updation. These expectation frameworks can be extremely useful in dynamic to deal with the current situation to control early intercessions.

This examination plans to give an early forecast model for the spread of novel corona virus, otherwise called SARS-CoV-2, and formally named as COVID-19 by the World Health Organization (WHO)[9].COVID-19 is directly an intense danger to human life everywhere throughout the world. Toward the finish of 2019, the infection was first recognized in a city of China called Wuhan, when countless individuals created indications like pneumonia [10].Its diverse affects is on the human body, including extreme intense respiratory disorder and multi-organ disorder which can eventually prompt demise in an exceptionally terms[11]. A huge number of individuals are influenced by this pandemic all through the world with each coming day. Many new individuals are in danger consistently from nations over the world. The main spread of infection essentially through close physical contacts, by respiratory channels, or by contacting the sullied surfaces. The testing part in this spread is that an individual can have the infection for a long time without indicating side effects. The reasons for its spread and thinking about its threat, practically all the nations have announced either severe lockdowns all through the influenced locales and urban communities. Clinical analysts all through the globe are now required to find a suitable immunization and drugs for the sickness. Since there is no endorsed prescription till now for executing the infection so the legislatures of all nations are concentrating on the safety measures which can stop the spread. Out everything being equal or either be educated, pretty much all the parts of COVID-19 are considered critical. To add to this part of data, various analysts are considering the various elements of the pandemic and produce the outcomes to support mankind.

To add to the current human emergency our endeavor in this examination is to build up an anticipating framework for COVID-19. The three significant parameters are considered to determine the effect of the pandemic for the coming 10 days:

-
  The count of new affirmed cases.
-

-
  The count of death cases.
-
  The count of recovered cases.

This issue of forecasting has been considered as a relapse issue in this study, so the investigation depends on some administered ML models, for example, supervised as well as unsupervised Learning. The learning models have been prepared by utilizing the COVID-19 dataset from the accessible standard website i.e. Kaggle named as covid_19_complete_dataset.csv. The assessment has been completed and the process of machine learning is applied for the forecasting using the accessible dataset. This study has some key discoveries which are recorded underneath:

- Forecasting performs best when the time-arrangement dataset has exceptionally restricted sections.
- Distinctive ML calculations appear to perform better in various class forecasts.
- The advancement in the ML calculations requires an abundant measure of information to anticipate the future, as the size of the dataset expands the model exhibitions improve.
- ML model-based anticipating can be extremely helpful for decision making for pandemics like COVID-19.

The paper comprises of six segments. Segment I presents the Introduction, segment II contains the related work, segment III contains the portrayal of the dataset and strategies utilized in this examination, segment IV presents the results and discussion and segment V concludes the paper.

2. RELATED WORK

Yang et al. (2020) presented dynamic SEIR model for foreseeing the COVID-19 epidemic zeniths and sizes. They used an AI model arranged concerning past SARS dataset also shows ensure for future desire for the scourges. Barstugan et al. (2020) introduced beginning phase area of COVID-19, which is named by World Health Organization (WHO), by AI techniques completed on Computed Tomography pictures. Elmousalami and Hassanien (2020) presents a relationship of day level foreseeing models on COVID-19 influenced cases using time arrangement models and numerical enumerating. Rizk-Allah and Hassanien (2020) familiar another foreseeing model with inspect and check the CS of COVID19 for the coming days reliant on the reported information since 22 Jan 2020. Rezaee et al. (2020) presented a blend approach reliant on the Linguistic FMEA,FuzzyInference System and Fuzzy Data Envelopment Analysis model to find out a novel score for covering some RPN insufficiencies and the prioritization of HSE perils. Navares et al. (2018) presented a response for the issue of envisioning each day clinical facility confirmations in Madrid due to circulatory and respiratory cases subject to biometeorological markers. Sujatha and Chatterjee (2020) proposed a model that could be valuable to predict the spread of COVID-2019 by utilizing direct relapse, Multilayer perceptron and Vector auto-regressionmodel on the COVID-19 kaggle information to imagine the epidemiological case of the disease and pace of COVID-2019 cases in India. Cui and Singh (2017) madeand applied the MRE speculation for month to month stream flow forecast with spectral power as an irregular variable. Torky and Hassanien (2020) presented a block chain consolidated structure which research the opportunity of utilizing distributed, time venturing and decentralized capacity focal points of block chain to build another system for confirming and recognizing the dark polluted occasions of COVID-19 contamination. Ezzat and Ella (2020) a novel technique called GSADenseNet121-COVID-19 reliant on a cross breed CNN structure is proposed using an improvement methodology. There is a need to mind the progressing variety in the provisions of the covid19 affected individuals to dissect the expanding or diminishing patterns for the different Countries.

3. DATA-SET AND METHODOLOGY

The machine learning make use of the algorithm that comprises of an objective/result variable (or dependent variable) which is to be expected from a given arrangement of predictors (Independent Variables).Utilizing this arrangement of variables, we produce a function that map to the desired yields. The training proceeds until the model accomplishes an ideal degree of precision on that accessible training data. Another one is unsupervised learning where we don't have any objective or result variable to foresee/estimate. It is utilized for making clusters of the population in various groups, which is generally utilized for segmenting clients in various groups for explicit intercession.The methodology deals with the forecasting and there's a difference between the prediction & forecasting in terms of machine learning. Therefore,the Prediction utilizes the known samples and test model with the old samples. But as a case we just predicting as well as forecasting for the future days in order to show the rise/fall of the Covid19 cases with impact for update in the count that assist the administration to take early measures as possible.

The point of this investigation is the future anticipating of COVID-19 spread concentrating on the count of new confirmed cases with respect to the count of active cases, the count of death cases and the count of recovered cases using Supervised Learning. The dataset utilized in the investigation has been taken from the GitHub vault by the Centre to Systems Science and Engineering, Johns Hopkins University and ASCAPS Govt. measures for Covid 19 [12]. Dataset files are contained in the organizer of the GitHub archive named (covid_19_clean_complete.csv).The organizer contains day time-seriescollected in the form of tables including the count of confirmed,death cases, and recovered cases. All the information is collected from the day-to-day case report and the sample data are shown in figures 2, 3,4 respectively.

The Data-set follows the Machine Learning Flow that incorporates the training information, learning calculation, building of the model,performing the training and feedback. Prior to this, the original data-set needs to experience the information handling,data visualization and data wrangling that ideal aides in the acknowledged objectives to accomplish for the machine learning to be applied on the covid19 dataset.

3.1 METHODOLOGY

The Following Steps leads to accomplishment of the model for the Forecasting with the Original Data-set gathered from the verified vaults of kaggle by using Colab Notebook as observed in the fig.1(a).

- Read the CSV document.
- Collecting the data from df.info ().
- Parse the "Date".
- Drop () the Province/State comprising of NAN Values.
- Rename Column – {Country/Region} to {Country}.
- Analyse the Data for Finding "Dynamic Cases".
- Performing the Visualized Group By {Country} – {Confirmed, Deaths, Recovered, Active Cases}.
- Performing Forecasting utilizing "fbProphet"- quick and gives totally mechanized conjectures that can be tuned by hand by information.
- Building the Model with Confidence Level - 0.95%.
- Train the model-make the Future dates for forecasting.
- Consider the "ds", "ywhat ", "ywhat – lower " andywhat-upper" Values.
- Graphical portrayals for the various values for evaluated results.
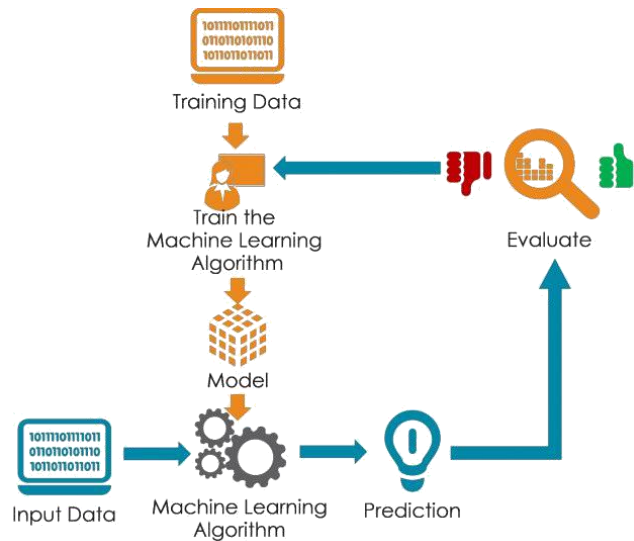
Figure 1 Illustrating the methodology for the Covid19 set

Before utilizing Python as a language, we must concentrate on to import the well predefined libraries utilized for better anticipating the outcomes that should visualise the fundamental forecasting results:

- Import pandas as pd
- Import seaborn as sns
- Import matplotlib.plyplot as plt
- Import plotply.express as px
- Import Prophet from fbprophet

Figure 2 Original dataset

The original dataset as in figure 2 incorporates the specific columns that may corelate with the better outcomes i.e. Province/state, Country/District and latitude, longitude, Date, Confirmed, Death and Recovered cases. Some way or another in machine learning the crude information needs to experience the procedure of data processing for the missing values,missing entities in order to manipulate the information in a sensible way.

The fig 3 incorporates the df.info () for the original dataset which assists in another degree of investigating the Null Worth Count and Dtype appeared for the accompanying type of data. Subsequent to data processing, the following stage is to change over the Dtype of "Date" in preferred date Format as appeared in Fig 4.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32065 entries, 0 to 32064
Data columns (total 8 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   Province/State  9680 non-null    object
 1   Country/Region  32065 non-null   object
 2   Lat             32065 non-null   float64
 3   Long            32065 non-null   float64
 4   Date            32065 non-null   object
 5   Confirmed       32065 non-null   int64
 6   Deaths          32065 non-null   int64
 7   Recovered       32065 non-null   int64
dtypes: float64(2), int64(3), object(3)
memory usage: 2.0+ MB
```

Figure 3 Information of Dataset

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered
32060	NaN	Sao Tome and Principe	0.186360	6.613081	2020-05-21	251	8	4
32061	NaN	Yemen	15.552727	48.516388	2020-05-21	197	33	0
32062	NaN	Comoros	-11.645500	43.333300	2020-05-21	34	1	8
32063	NaN	Tajikistan	38.861034	71.276093	2020-05-21	2350	44	0
32064	NaN	Lesotho	-29.609988	28.233608	2020-05-21	1	0	0

Figure 4 Converting Date into format

By one way or another in information processing, we discovered the Columns that needs to be dropped out for better outcomes.Therefore, the column named "Province/State" dropped so as to have normalised results as shown in figure 5.

	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered
0	Afghanistan	33.0000	65.0000	2020-01-22	0	0	0
1	Albania	41.1533	20.1683	2020-01-22	0	0	0
2	Algeria	28.0339	1.6596	2020-01-22	0	0	0
3	Andorra	42.5063	1.5218	2020-01-22	0	0	0
4	Angola	-11.2027	17.8739	2020-01-22	0	0	0

Figure 5 Dropping the Province/State columns

	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered	Active
32060	Sao Tome and Principe	0.186360	6.613081	2020-05-21	251	8	4	239
32061	Yemen	15.552727	48.516388	2020-05-21	197	33	0	164
32062	Comoros	-11.645500	43.333300	2020-05-21	34	1	8	25
32063	Tajikistan	38.861034	71.276093	2020-05-21	2350	44	0	2306
32064	Lesotho	-29.609988	28.233608	2020-05-21	1	0	0	1

Figure 6 Adding the Active Cases

Visualization of the data plays out important role as wrangling appeared in figure 6 to gauge the "Active Cases" extracted by using the Confirmed, Death and Recovered Cases that will be utilized in further

3155

anticipating to investigate the concerned covid 19 data.The remark of the increase / decrease count of the distinctive classified cases for the different geological areas under the column named "Country" shown in figure 7.

Country	Confirmed	Deaths	Recovered	Active
Afghanistan	8676	193	938	7545
Albania	969	31	771	167
Algeria	7728	575	4062	3091
Andorra	762	51	639	72
Angola	58	3	17	38

Figure 7 Grouping Data as per Data

The Figure 8 Shows the "Active cases" for topographical zones tended to with the hues recognizes extending from 0K to 100 K with the library. Making use of choropleth maps requires two primary kinds of information:

I. Geometry data:
- This can either be a provided GeoJSON document where each element has either an id field or recognizing by an incentive in properties; or
- one of the implicit geometries inside plotly: US states and world nations (see beneath)

II. A rundown of features listed by indexed identifier.

The GeoJSON information is passed to the geojson contention, and the information is passed into the shading contention of px.choropleth (z if utilizing graph_objects), in a similar request as the IDs are passed into the area arguments.
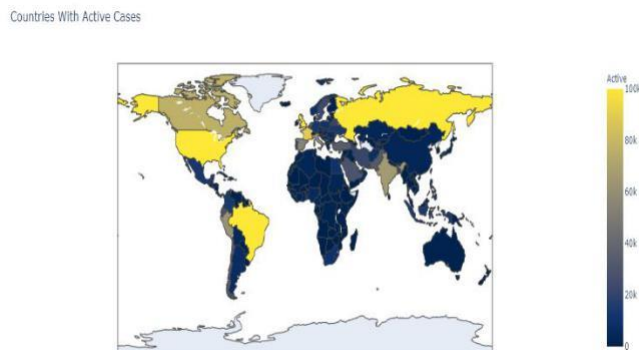

Figure 8 Active cases

The forecasting can be accomplished with the library named "fbProphet" that executes a methodology for anticipating time –series information that is dependent on an added substance model where non-straight patterns are fit with yearly, week by week, and day by day irregularity, in addition to procedural impacts. It works best with time –series that have great impacts and a few recorded data. Prophet is robust to missing information and movements in the pattern, and normally handles exceptions well. This will be seen in the segment V with the results and discussion that accomplished our future dates with the better

outcomes in order to update us with the upcoming trends of geological countries in terms of distinct parameters of case in nearer future of 80 days.

4. RESULTS AND DISCUSSIONS

As far as the discussion is concerned and the particular outputs achieved through the machine learning, this investigation endeavours to build up a framework for the future determining of the number of cases influenced by COVID-19 utilizing AI techniques. The LASSO as Least absolute shrinkage & selection operator being an regression analysis method is used as the technology to enhance the prediction accuracy and interpretability of the resulting statistical model. The dataset utilized for the investigation contains data for the different days and involve in it the count of recently affected cases,count of recovered Cases, and count of death cases due to COVID-19 around the world.As the demise rate and confirmed cases are expanding step by step, which is a disturbing circumstance for the world? The number of individuals who can be influenced by the COVID-19 pandemic in various nations of the world isn't notable. Thisinvestigation is an endeavour to forecasting for the count of individuals that can be influenced as far as new tainted cases and deaths including the count of anticipated recovered cases for the upcoming 80 days.

The green line depicts the normal general affirmed Cases i.e. Confirmed cases as the days are expanding step by step with the predicted model as shown in figure 9, "Total Cases vs. Days" from January to May. In rest of the outcomes we will discover the yields for the forecasted one.
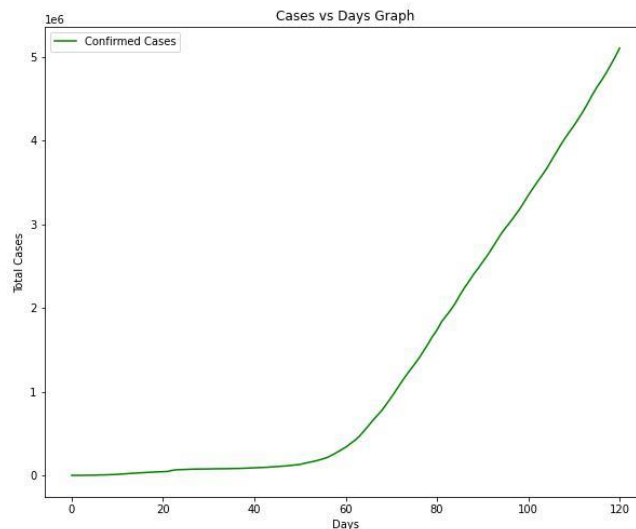
Figure 9 Confirmed Cases Vs Days Graph

The scatter bend lines of the Green, Blue, Purple and red lines portray the genuine cases as gathered from the Data-set for the dates of the long stretch of January - May which remembers the more increment for the confirmed cases relating to the less recovered cases as shown in figure 10. Here the overlapped point's shows the accuracy in the outcomes acquired.In the beginning, there shows a static extending and with the days it shows an enormous increment which influences the world's economy to tumble down.
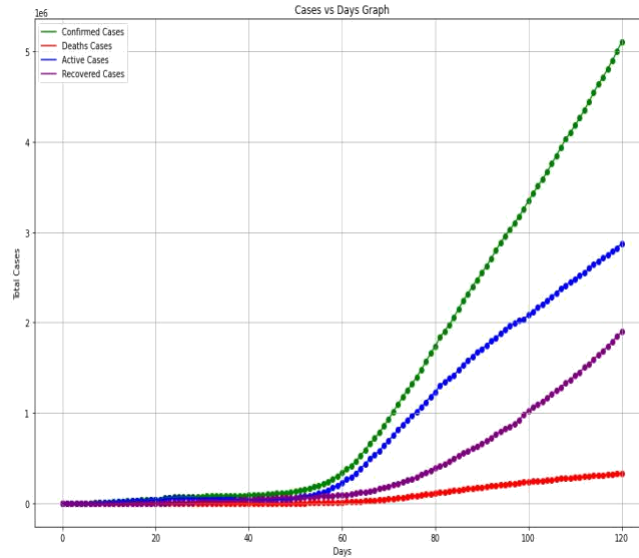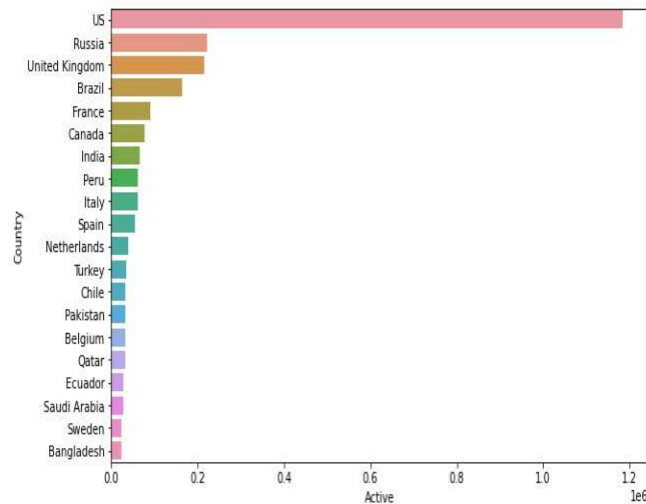
Figure 10 Scatter & Line Graph for Confirmed, Deaths, active and Recovered Cases from Jan – May

Figure 11 Top 20 Countries -Active Cases

**Figure 12** Bar Plot for Active Cases

The figure 11 shows the tally of the top20 dynamic cases and the figure 12 shows the Bar plot to graphically investigate the outcomes in progressively legitimate manner.According to the Data-Set the predicted value shows the United States with the more increment in the Count of Active cases in contrast with the Russia for the period of Jan-May with the difference of 962,253. India positioned at the seventh situation as per the Data-Set with the difference of 1,117,938 in contrast with the United States with driving situation in the grievous run of Covid-19 cases. The Figure 13 shows the graphical portrayal with the hued bends in comparison with the India, US, China and Germany with the x-pivot as Days and Y-pivot as Deaths.
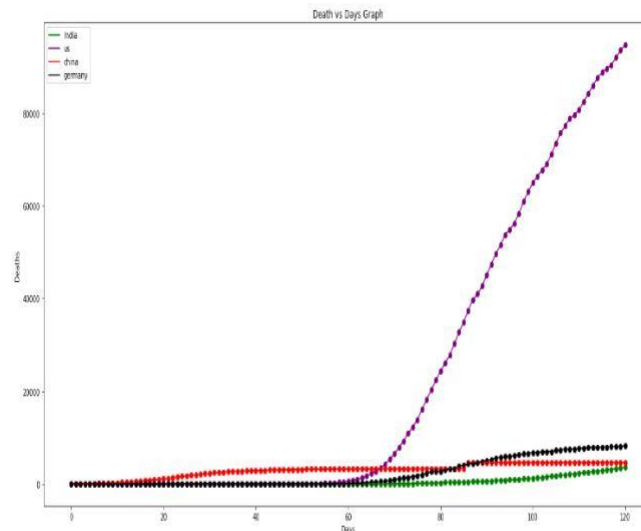
**Figure 13** Scatter & line Curves for Death Vs Days Cases- India, US, China & Germany

Quite a bit of AI includes evaluating the performance of an AI calculation on inconspicuous information. Confidence Intervals are a method of evaluating the vulnerability of an estimate. They can be utilized to include a limits or probability a populace boundary, for example, a mean, assessed from a sample of observations from the population. Confidence interval originates from the field of estimation insights. With the small sample of Data-set, the 95% certainty span is like the scope of the information. Be that as it may, just a little portion of the values in the enormous sample on the correct exist in the confidence level. The 95% certainty stretch characterizes a scope of values that you can be 95% sure contains the mean of the Population. With huge samples, you realize that mean with significantly more accuracy than you do with a little sample, so the confidence level is very thin when processed from an enormous sample of data-set. The future dates are framed for 80 days starting from 22nd, May 2020 in sequence to the forecasting for the later on days.

The model is being built with the processed data-set having the confidence level i.e. interval-width 0.95%. As per figure 14 is concerned we renamed the two columns i.e. Date and Confirmed Cases as "ds" & "y respectively. For Forecasting the predicted value, Lower range and upper range values signified by the "yhat", "ywhat_Lower" and "ywhat_Upper respectively.
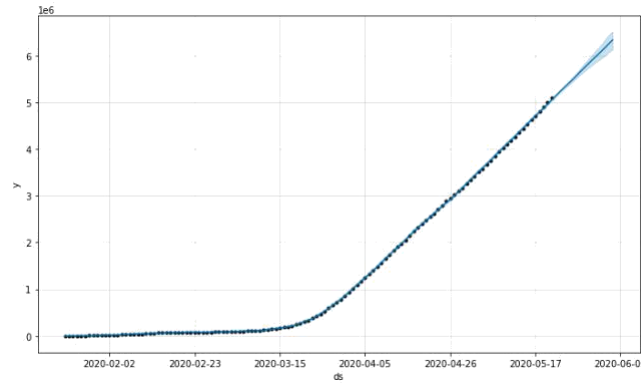
Figure 14 Forecasting: Date" ds" vs. Confirmed Cases "y" forecasting

The figure 15 shows the breakage in the stretch out of chart from May onwards to the accompanying 80 days for which we are estimating, which unmistakably means a drop in the Death Cases as between the scope of 2000 – 4000 and same between the 4000 – 6000. The comparable measurements demarked the genuine Death rates as determined in the " Covid19 ready Statistics".
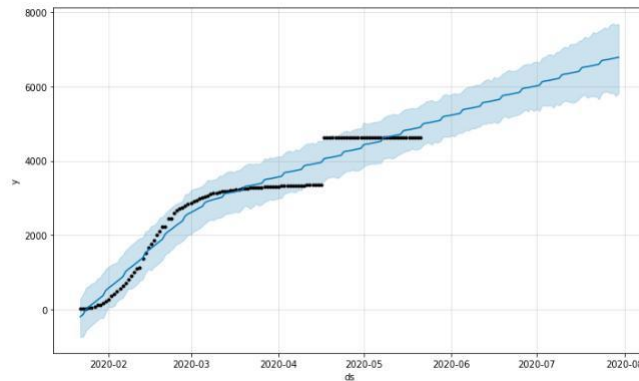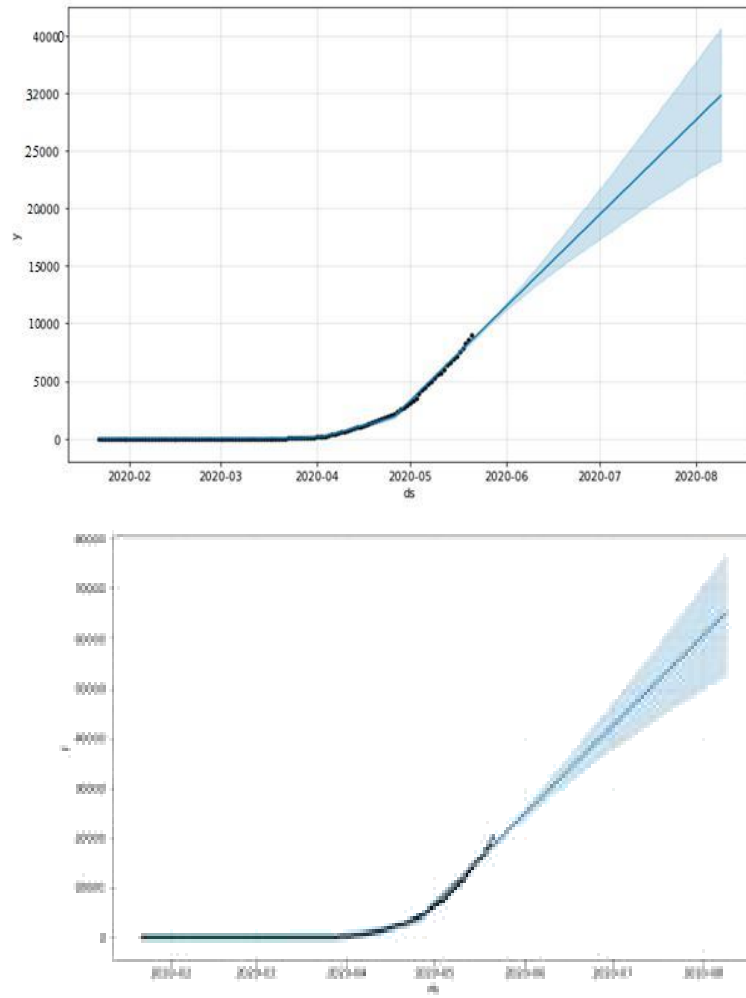

Figure 15 Forecasting the Death Cases for "China"

The significant outcomes for the Death instances of the United States can be accomplished as appeared in figure 18 where the concealed part shows the tremendous forecasting determining of the bends that can be found in a manner for the future dates of succeeding 80 days using LASSO and therefore the real passing insights for covid19 in US is very nearly 149 K that simply going to influence the world economy and we need to take prompt action. Likewise, the development of Brazil demise rate expanding as the graph shows in figure 19 inside the extents over 8000 and may show huge increment in closer future moreover.


Figure 16 Statistical Covid19 Confirmed, Recovered and Death cases.

Figure 17 Forecasting the Death Cases for "India"

In figure 17, the outcomes can be observed for India – Forecasting for the death Cases where the "ywhat _Lower extend" i.e. 20,000 and the "ywhat_Upper extend" i.e. 40,000 which can implicit appears to increment in closer future too. In many cases, the actual observed outcomes for the death cases are more i.e.32,771 till July and the graph shows that the informational index for the Covid19 is marginally demonstrating the significant result with the faded shade part in comparsion to the exponential smootheing.
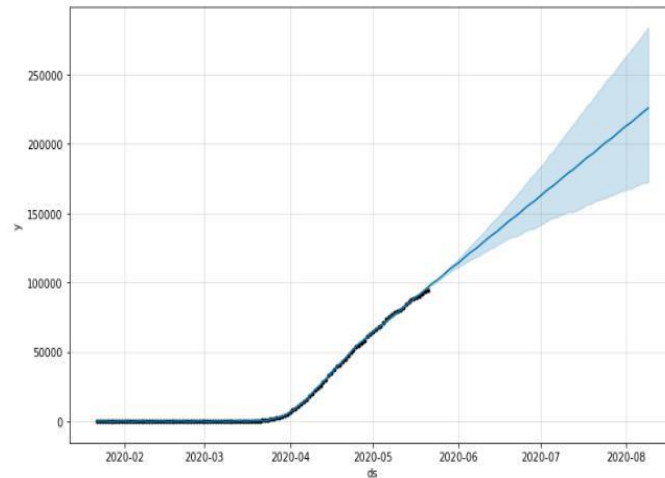
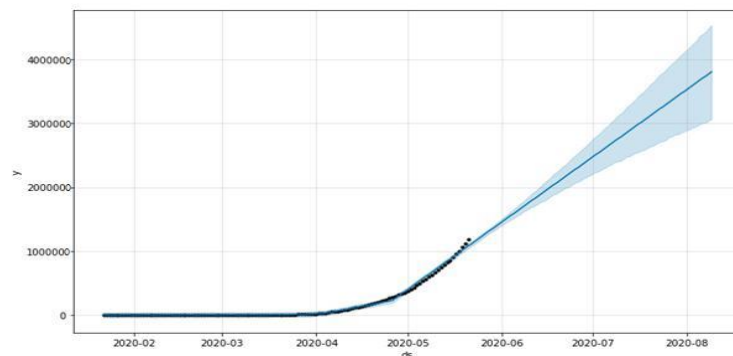**Figure 18** Forecasting the Death Cases for "US"

The significant outcomes for the death instances of the United States can be accomplished as appeared in figure 18 where the concealed part shows the tremendous forecasting determining of the bends that can be found in a manner for the future dates of succeeding 80 days and therefore the real passing insights for covid19 in US is very nearly 149 K that simply going to influence the world economy and we need to take prompt action. Likewise, the development of Brazil demise rate expanding as the graph shown in figure 19 inside extends over 8000 and may show huge increment in closer future moreover.

Figure 19 Forecasting the Death Cases for "Brazil"

3162
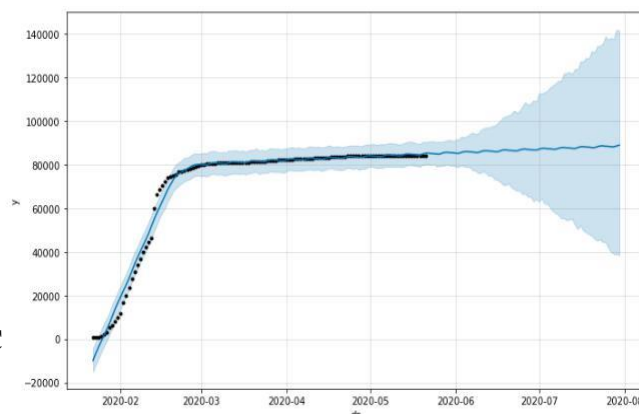
Figure 20 Forecasting the Confirmed Cases for "China"

These comparable graphs as seen in the figure 20 with likely increment in the affirmed cases for the China that shows the consistent bend between the slopes of 8000 – 10,000 and matches with the genuine measurements of Covid19 affirmed cases. The affirmed cases as from the graph portrays shows no variety in the long stretch of February and as the days goes on huge up-lift has been appeared by the graph in the development of Confirmed cases which influences the human life and the government imposed the consecutive lockdown in the regions of India with effect from March yet at the same time the cause increases the development which came across to 14, 35,000.From the graph the estimating likewise shows the expansion in the closer 80 days beginning from the long stretch of May.

Figure 21 Forecasting the Confirmed Cases for "India"
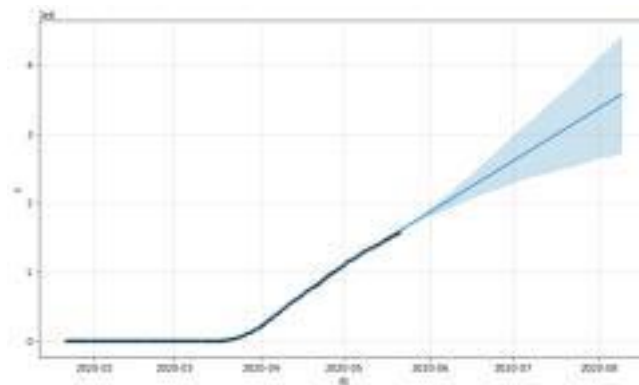
Figure 22 Forecasting the Confirmed Cases for "US"

The Forecasting development of Confirmed cases shows in the long stretch of June- August provides us the exponential values and may increment in closer days as observed in figure 22. The estimation drawn is similar to the standard covid19 insights from authenticated accessible sources which is liable to have the most extreme values.When contrasted with other countries in figure 20, 21 and 23 the forecasting growth rate in figure 23 likewise shows the expanding standards and furnishes the again exponential development with the legend esteems in the long stretch of June-August as anticipated. But as observed in February the rate is consistent and shows no impact of covid19 in that topographical locale.

5. CONCLUSIONS

According to WHO, the forecasting rate in the US reached to 43,12,841 which is likely shown by the graphs drawn with the data-set from the revoked sources with the assistance of AI training and building of the model in new period of the today's era. In this investigation, a ML-based expectation framework has been proposed for estimating the danger of COVID-19 pandemic comprehensively.The model analyses the dataset that contains the day-wise real past information with different investigation of reports from 20 nations, including India, put the more predicted one outcomes as observed in above graphs with progressively scientific view for evaluating and new measures to be taken by the administration utilizing the most recent advancements for humankind and it generally brings about new variation in the Confirmed, Death and Recovered instances of Covid19 in order to make vital move ahead.The result shows that in more closer days great exponential development is observed in the country of United States in comparison to the rest of the Countries as indicated by the processed data-set, for which the forecasting is implemented through machine learning technique i.e. LASSO and Exponential smoothening .From the above data , LASSO is much stable as comparion to the exponential smoothening as it reduces the error in some extent as shown by the results and discussion segment . In real pattern, we will compare the countries boundaries with the other countries and predict the impact of Covid19 in larger extent in future work.

REFERENCES

[1] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. (2020). doi: 10.1016/S1473-3099(20)30120-1. [Epub ahead of print].

[2] Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? Lancet. (2020). doi: 10.1016/S0140-6736(20)30627-9. [Epub ahead of print].

[3] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N Engl J Med. (2020) 382:1199–207. doi: 10.1056/NEJMoa20 01316

[4] Zhou T, Liu Q, Yang Z, Liao J, Yang K, Bai W, et al. Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. J Evid Based Med. (2020) 13:3–7. doi: 10.1111/jeb m.12376

[5] Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. Int J Infect Dis. (2020) 92:214–7. doi: 10.1016/j.ijid.202 0.01.050

[6] Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med. (2020) 27:taaa021. doi: 10.1093/jtm/taaa021

[7] Lai A, Bergna A, Acciarri C, Galli M, Zehender G. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. J Med Virol. (2020). doi: 10.1002/jmv.25723. [Epub ahead of print].

[8] Ippolito G, Hui DS, Ntoumi F, Maeurer M, Zumla A. Toning down the 2019nCoV media hype – and restoring hope. Lancet Respir Med. (2020) 8:230–1. doi: 10.1016/S2213-2600(20)30070-9

[9] McCloskey B, Zumla A, Ippolito G, Blumberg L, Arbon P, Cicero A, et al. Mass gathering events and reducing further global spread of COVID19: a political and public health dilemma. Lancet. (2020) 395:1096–9. doi: 10.1016/S0140-6736(20)30681-4

[10] Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. Rev Mod Phys. (2015) 87:925. doi: 10.1103/RevModPhys.87.925

[11] Wang Z, Bauch CT, Bhattacharyya S, d'Onofrio A, Manfredi P, Perc M, et al. Statistical physics of vaccination. Phys Rep. (2016) 664:1–113. doi: 10.1016/j.physrep.2016.10.006

[12] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D. Complex networks: structure and dynamics. Phys Rep. (2006) 424:175–308. doi: 10.1016/j.physrep.2005.10.009

[13] Holme P, Saramäki J. Temporal networks. Phys Rep. (2012) 519:97–125. doi: 10.1016/j.physrep.2012.03.001

[14] Boccaletti S, Bianconi G, Criado R, del Genio C, Gómez-Gardeñes J, Romance M, et al. The structure and dynamics of multilayer networks. Phys Rep. (2014) 544:1–122. doi: 10.1016/j.physrep.201 4.07.001

[15] Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer networks. J Complex Netw. (2014) 2:203–71. doi: 10.1093/comnet/cnu016

[16] Lü L, Chen D, Ren X-L, Zhang Q-M, Zhang Y-C, Zhou T. Vital nodes identification in complex networks. Phys Rep. (2016) 650:1–63. doi: 10.1016/j.physrep.2016.06.007

[17] Salathe M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoSComput Biol. (2012) 8:e1002616. doi: 10.1371/journal.pcbi.1002616

[18] Helbing D, Brockmann D, Chadefaux T, Donnay K, Blanke U, WoolleyMeza O, et al. Saving human lives: what complexity science and information systems can contribute. J Stat Phys. (2015) 158:735–81. doi: 10.1007/s10955-014-1024-9

[19] Perc M. The social physics collective. Sci Rep. (2019) 9:16549. doi: 10.1038/s41598-019-53300-4

[20] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. (2020) 395:497–506. doi: 10.1016/S0140-6736(20) 30183-5

[21] Lazzerini M, Putoto G. COVID-19 in Italy: momentous decisions and many uncertainties. Lancet Glob Health. (2020). doi: 10.1016/S2214-109X(20)30110-8. [Epub ahead of print].

[22]   Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following covid-19 infection. Lancet Infect Dis. (2020). doi: 10.1016/S1473-3099(20)30195-X. [Epub ahead of print].

[23]   Chen D, Xu W, Lei Z, Huang Z, Liu J, Gao Z, et al. Recurrence of positive sars-cov-2 rna in covid-19: a case report. Int J Infect Dis. (2020). doi: 10.1016/j.ijid.2020.03.003. [Epub ahead of print].

[24]   Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. (2020) 579:270–3. doi: 10.1038/s41586-020-2012-7

[25]   Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. J Med Virol. (2020) 92:522–8. doi: 10.1002/jmv.25700

[26]   Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2. Nat Sci Rev. (2020) nwaa036. doi: 10.1093/nsr/ nwaa036

[27]   Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med. (2020). doi: 10.7326/M20-0504.

[28]   Johns Hopkins University Data Repository. Cssegisanddata. Accessed: Mar. 25, 2020. [Online]. Available: https://github.com/ CSSEGISandData.

[29]   S. Baran and D. Nemoda, ''Censored and shifted gamma distribution based EMOS model for probabilistic quantitative precipitation forecasting,'' Environmetrics, vol. 27, no. 5, pp. 280–292, Aug. 2016.

[30]   Y. Grushka-Cockayne and V. R. R. Jose, ''Combining prediction intervals in the m4 competition,'' Int. J. Forecasting, vol. 36, no. 1, pp. 178–185, Jan. 2020.

[31]   N.C.Mediaite.HarvardProfessorSoundsAlarmon'Likely'Coronavirus Pandemic: 40% to 70% of World Could be Infected This Year. Accessed: Feb. 18, 2020. [Online]. Available: https://www.mediaite.com/news/ harvard-professor-sounds-alarm-on-likely-%coronavirus-pandemic-40to-70-of-world-could-be-infected-this-year/

[32]   BBC. Coronavirus: Up to 70% of Germany Could Become Infected— Merkel. Accessed: Mar. 15, 2020. [Online]. Available: https://www.bbc. com/news/world-us-canada-51835856