International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

Statistical Analysis of COVID-19

A. Dubey[1], A. Sharma[2],P. Tiwari[3], V. Kate[4]

[1,2,3]*Student, Department of Computer Sciences and Engineering, Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India*

[4]*Associate Professor,* Computer Sciences and Engineering, Acropolis Institute of Technology and Research, Indore, Madhya Pradesh, India

---

Abstract -The recent outbreak of COVID-19 Coronavirus has brought forth a very cataclysmic situation globally and has lead this world to a standstill. As the virus encroaches further into our domain, it has become crucial to monitor its growth and keep the ones compromised in check. In this study, we are striving to fuse Machine Learning and Data Science in order to get an insight into the problem. In this study Prophet Algorithm is used, which is stated to be accurate in these scenarios. Lastly, analysis of its accuracy is also conducted for the present scenario.

Key Words:Prophet Algorithm, Machine learning, MAPE, COVID-19, Data Science.

1. INTRODUCTION

On December 31st 2019, China intimated WHO regarding the peculiar cases of pneumonia, for which a new term was coined i.e. "COVID-19", followed by the very first demise registered by it on February 7th. By the fall of February 2020, COVID-19 was declared a pandemic by the World Health Organization. Since then, there has been a rapid increase in the number of casualties as far as the first wave is concerned. The primary reason for this can be traced to the highly dynamic nature of the virus which in turn enables it to mutate and increase its structural complexity. Thus, it has become the need of the hour to eradicate this bane before it delivers a catastrophic blow to mankind [1].

2. Body of Paper

The Prophet is an open-source library published by the developers at Facebook which is predominantly based on decomposable models. It is a time-based pattern forecasting model which possesses the ability to make precise deductions through its dual-staged process and simple, intuitive parameters.

Furthermore, it is plausible to verify the accuracy of the forecast model by matching the deduced results with the ones produced earlier through MAPE. Due to this, the risk of producing erroneous results plummets drastically, therefore conferring Prophet, an upper hand over the other algorithms. Research has shown that the chance of a correct diagnosis in difficult cases is 5.8% and worse and that the doctors involved are 64% certain about the inferred outcomes. Hence, predicting the future in the case of any pandemic or endemic turns out to be a laborious task due to the copious amounts of data sets. Despite this, we sincerely hope that this project serves as a tool of great assistance for the analysts, to observe the growth trends of the active cases of COVID-19 demographically and thus, take required precautionary measures. (Who are relentlessly working with the objective of fortifying our world against this calamity, in anticipation of a better tomorrow). In this study, we provide statistical estimates of the confirmed cases of COVID-19 using time series analysis, which can yield more accurate results than regression algorithms. [2]

This paper includes the follow up of the Prophet algorithm with the objective of providing an insight into the fluctuating growth trends of the COVID-19 for the 2nd wave. The rest of the work can be organized as follows; SECTION 2 – Concept of Prophet, Section 3 Performance analysis, Section 4 Conclusion.

Concept of Prophet

The Prophet is an open-source forecasting tool available in Python and R, which has been developed by Facebook. It, at its core, is an additive regression model with four main components.

A piecewise linear or logistic growth curve trend , a yearly Fourier series, a weekly seasonal component using dummy variables, and a user-provided list of important holidays.[4]

Prophet automatically detects changes in trends by selecting change points from the data. [1]

In its simplest form:

$$y(t) = g(t) + s(t) + h(t) + e(t)$$

where:

g(t): trend models non-periodic changes(growth over time)

h(t) : It ties in effects of holidays.

s(t) : seasonality presents periodic changes

e(t) : covers idiosyncratic change not accommodated by the model

The equations can also be written as:

y(t) = piecewise_trend(t) + seasonality(t) + holiday_effects(t) + noise[1]

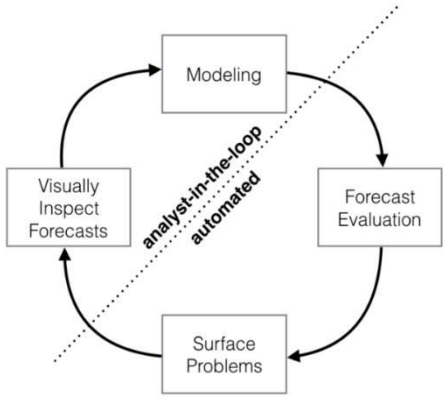

Fig -1: Prophet Working model [1]

3. Trend Analysis and Forecast

This study is focused on experimenting and testing the Prophet algorithm with the offline data of the COVID-19 accumulated through various source lines and attempting to predict the growth rate of the victims for the upcoming days. We have scrutinized the performance of the algorithm, since it possesses an established reputation when it comes to accuracy. Furthermore, the dexterity of Prophet has been proven when it comes to producing results that can be comprehended by any individual irrespective of their expertise in the field. It can greatly assist the analysts to distinguish among the areas which are more prone to further spread and have a comparatively higher degree of COVID-19 patients to the insignificant ones and thus, alerting the people and authorities concerned to take required precautionary measures post apprehending the associated risks.

In this study, we have observed the rate of growth of the Coronavirus in different countries, by plotting various graphs, to deduce various parameters which can determine a trend, so that, assumptions can be made regarding the upcoming cases worldwide. A zone map of India is plotted where the red zones represent the most contagious areas. This testing was carried out two times, once for March and the other for the month of June-August.

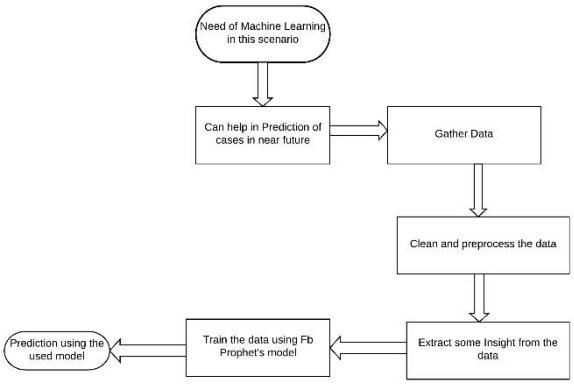

Fig -2: Flowchart

4. Forecasting COVID-19

Also, we will work on forecasting COVID-19 transmission. This model follows the sklearn model API. We have devised an instance of the Prophet class and then used fit to train the model and predict the method to obtain further outcomes. The input of the model is a data frame with two columns: ds and y. The ds (date stamp) column should be of a format expected by Pandas, ideally YYYY-MM-DD for a date or YYYY-MM-DD HH: MM: SS for a timestamp. The y column must be numeric and represents the measurement scale we wish to forecast. Now we will use the input ds and y as data in our algorithm and then build a data frame on which we will get daily predictions for the next month. [3]

Below graph clearly shows exponential growth in the positive cases of Coronavirus in India cases going over 800000 in the month of June-July.

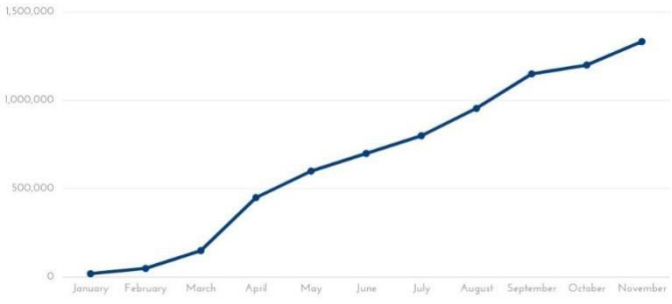

Fig- 3: (datastampvs positive cases)

A. Forecast Quality Evaluation

In this section, we have tested the accuracy of the model for different sets of data. Determination of accuracy of the model is achieved by performing cross-validation on the pevious data. Here, we implement the data of the last four months to deliver a forecast of data for the next month.

As per the preferences of Facebook's team, the author of the Prophet Paper, absolute percentage error (MAPE) comes out to be the best criterion to check the accuracy of the model.[4]

Performance Matrix -

Horizon is the time interval taken into consideration, the main aspect of determining the accuracy is MAPE which should be less than that of 10% mark, which seems to be well maintained in the results.

```
[97] from fbprophet.diagnostics import performance_metrics
     df_p = performance_metrics(df_cv)
     df_p.head()
```

	horizon	mse	rmse	mae	mape	mdape	coverage
0	5 days	9.104756e+05	954.188474	851.310557	0.667979	0.648179	0.0
1	6 days	1.798561e+06	1341.104508	1208.581389	0.719024	0.727675	0.0
2	7 days	3.157515e+06	1776.939874	1634.152377	0.758158	0.777500	0.0
3	8 days	5.442083e+06	2332.827215	2179.922056	0.803892	0.816534	0.0
4	9 days	8.649762e+06	2941.047751	2793.191791	0.835639	0.820905	0.0

Fig -4: Performance Matrix

Cross-Validation-

The initial dataset is split into separate training and testing data subsets followed by holding out a set of data and then, training the model accordingly in conjunction with the rest of the data. Hereafter, the model on the hold dataset is tested and the evaluation score of the model is retained and the procedure is recurred to obtain the optimized fitting model.

In the graph below the trend line shows the predicted values and the actual values are shown by marks plotted on the graph. It clearly shows that there is a very minute error of the model between the trend line and the actual outcomes which represent the accuracy of the model [5]

```
[99] from fbprophet.plot import plot_cross_validation_metric
     fig = plot_cross_validation_metric(df_cv,metric='mape')
```

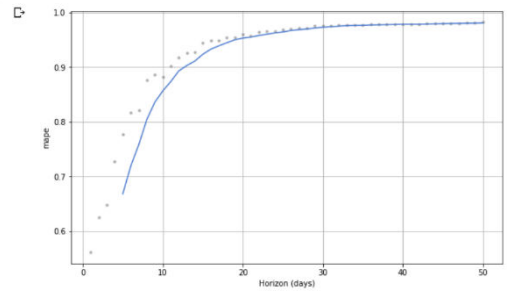

Fig-5: Cross Validation

5. CONCLUSIONS

The main objective of this study was to analyze the fluctuating growth in the number of individuals compromised due to COVID-19 and then assessing our observations with respect to Prophet Algorithm. After taking the model's compatibility and the results produced into account, we have observed that, the model fits with an accuracy of 91%. As of 5th April 2021, India reported a total of 12,589,067 confirmed cases and 103,558 confirmed cases within the past 24 hours. This is highest number of daily COVID-19 cases reported by India. A total of 165,101 deaths have been reported. Through our research, we've concluded that the cases of corona will increase exponentially unless the administered vaccine takes effect. Since, the virus is on a rampage, social policies need to reflect a commitment to respect and protect the most vulnerable individuals and to create the conditions for the fulfillment of economic and social rights for all. We believe that the claims made will significantly contribute in embracing and fortifying us for the future.[6][7][8]

6. REFERENCES

[1] Corona Virus Resource Centre, John Hopins University

[2] Meyer AN, Payne VL, Meeks DW, Rao R, Singh H. Physicians' diagnostic accuracy, confidence, and resource requests: A vignette study. JAMA Intern Med. 2013; 173:1952–8.

[3]Senthil E, 2019, Apple Stock and Bitcoin Price Predictions Using FB's Prophet-Using Python

[4] Upasana, May 2020, COVID-19 Outbreak Prediction using Machine Learning | Edureka.

[5] Sean Taylor [cre, aut], Ben Letham [aut], Automatic Forecasting Procedure, Quick Start | Prophet by Facebook

[6] COVID-19 status, Government of India https://www.mygov.in/covid-19

[7]Coronavirus-disease-(covid-19)-India-situation-report: 62nd WHO (issued by)

[8]Maritza Vasquez Reyes- STUDENT ESSAY The Disproportional Impact of COVID-19 [Link]