

COVID-19 Future Forecasting based on Time-series statistical analysis using Machine Learning Model

Jaspreet Kaur¹, Prabhpreet Kaur²

^{1,2}Department of Computer Engineering & Technology, Guru Nanak Dev University, India

¹jaspreet.kaur144@gmail.com

²prabhpreet.cst@gndu.ac.in

Abstract

The epidemic COVID-19 has shaken the globe through its cruelty, and its spread rate continues to rise daily. This paper highlights the clinical stance in the COVID-19 research studies, where time-series statistical analysis has been performed by using Prophet Model. It is widely used to understand the trend of the current epidemic after 2nd May 2020 with data at the worldwide state. The prophet model is an open-source model obtained by the data science panel on Facebook for performing predicting operations. It assists to make fast and accurate predictions for existing data samples. The Prophet model is simple to implement because its open authorized repository exists on GitHub. The time-series data analysis refers to the confirmed, recovered, and death rates for the time of 2nd May 2021 to 17th January 2022. The statistical validation strategy is followed by the implementation of a T-test on the evaluated time-series data. The expected data generated by the predictive model can be further used by the official authorities, medical departments of various countries. Moreover, the model is used to provide new graphical insights into past, present, and future trends.

Keywords

COVID-19, Prophet, Statistical analysis, time-series forecasting system.

1. Introduction

The coronavirus emerged from Wuhan city, China in December 2019. Originally, an unidentified infected case was reported which was examined by respiratory experts affirmed as pneumonia. Afterward, it was stated by WHO (World Health Organization) as COVID-19 [1]. It is at the seventh number of the coronavirus family, collectively with MERS (Middle-East Respiratory Syndrome) [2] and SARS (Severe Acute Respiratory Syndrome) [3] which can transmit to humans [4]. The spreading rate of the coronavirus rapidly increases worldwide. There are 236,624,144 confirmed cases, 213,744,952 recovered cases, and 4,832,164 death cases globally on 6th October 2021 [5]. The COVID-19 affecting individuals worldwide with mild to moderate symptoms as cough, fever, and fatigue. The infection can cause serious medical problems such as heart complications, lung infections, blood clotting, and severe kidney damage, bacterial and viral infection. The incubation phase of COVID-19 can proceed for 2 weeks or be extended. The data showed that the virus gets widespread from one individual to another in a limit of six feet or two meters [6]. Presently, the governments are considering preventive measures like sanitization, social distancing, strict lockdowns, etc. The study employ with time-series-based statistical framework model as “Prophet”, which shown the accurate results in forecasting both short and long-term prediction measures. It allows evaluating the considerable trends, seasonality, cyclic effect, and abnormality measures. The main purpose of the article is to represent the approximate 8.5 months (260 days) predict the trends

International Conference on Smart Systems and Advanced Computing (Syscom-2021), December25–26, 2021

EMAIL: jaspreet.kaur144@gmail.com (A. 1); prabhpreet.cst@gndu.ac.in (A. 2)

ORCID: 0000-0002-2462-5182 (A. 1); 0000-0001-8498-5940 (A. 2)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

for confirmed, recovered, and death cases for different countries. **Figure 1** determines the map of countries showing confirmed cases from the sample dataset.

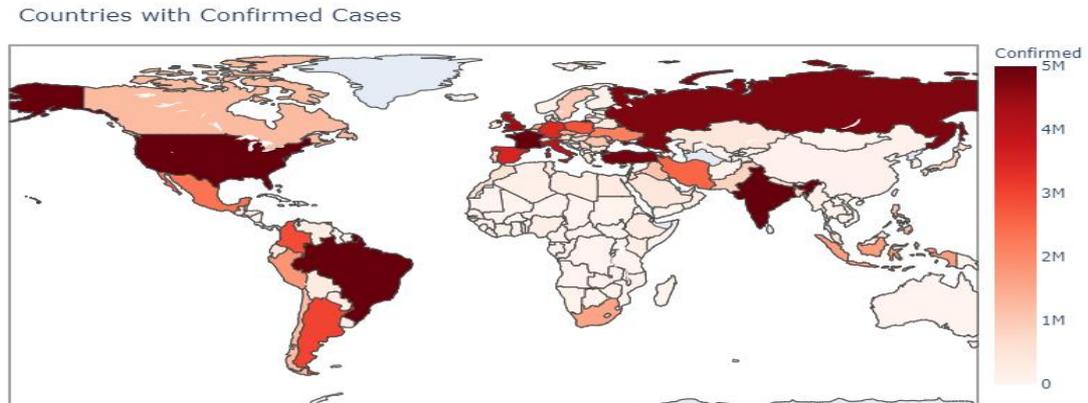


Figure 1: World map - Countries with Confirmed cases

2. Literature Survey

Anastassopoulou et al. [7] performed the experimental analysis on the COVID-19 data by using the SIDR (Susceptible-Infectious-Recovered-Dead) framework. This model provides approximations of the basic infectious reproduction number (R_0). It is used to predict the growth rate for up to three weeks. **Alsaedy et al.** [8] conducted work based on recognition of areas, which are highly prone to spreading COVID-19 utilizing a wireless network. In this study, they employed an end-user machine (UM) that connected with a wireless cellular network mechanism for better inference of regions. **M. B. Jamshidi et al.** [9] conducted a DL (Deep learning) approach for the detection of COVID-19 infection. They performed AI-based strategies to diagnose the coronavirus infection by using GAN (Generalized Adversarial Network), LSTM (Long Short-Term Memory), and ELM (Extreme Learning Machine). **AL-Rousan et al.** [10] conducted the COVID-19 analysis, which describes the exponential growth of the infected cases in South Korea. **Lutz et al.** [11] conducted the infectious disease forecasting framework, using various mathematical models. In this study, the research has been done on human beings and the way how it interrelated with infectious ailments with handling approaches. **Guo et.al** [12] performed a forecasting model by using Prophet for (MPD) Maximum-Power Demand along with adaptive Kalman filter. **Chae et al.** [13] conducted a forecasting system based on deep learning and a big data approach for time analysis of COVID-19. In **Figure 2**, the diagrammatical representation of the COVID-19 prediction framework for the next 260 days.

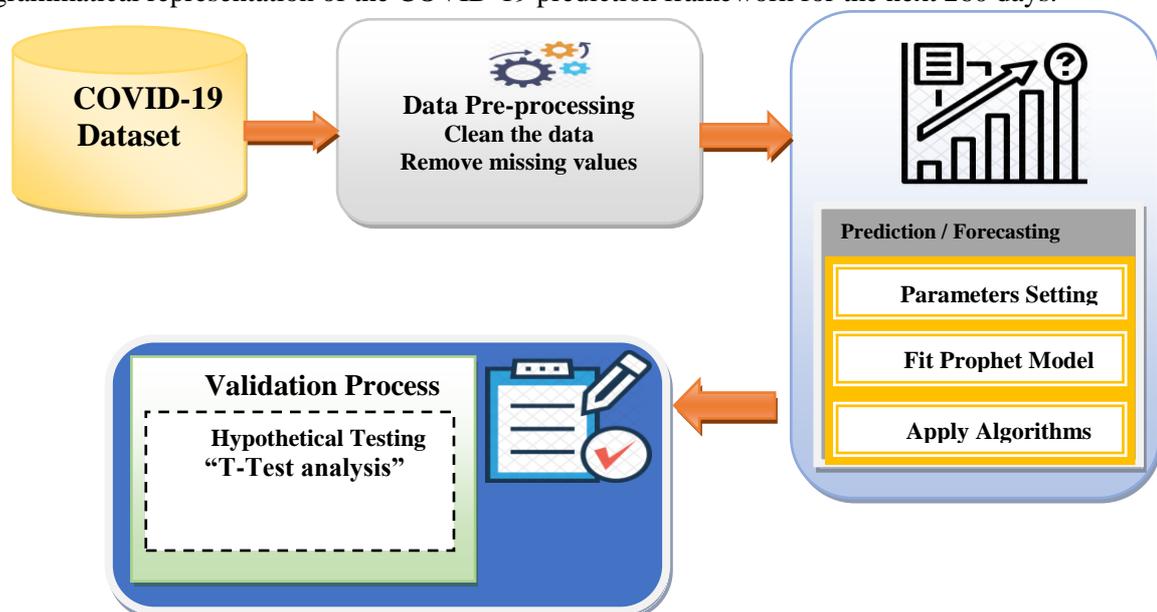


Figure 2: Schematic representation of COVID-19 prediction model with statistical analysis.

3. Data Description and Materials

The Prophet framework is a practice for time-series-based forecasting of the COVID-19 data. It is depending on the additive modeling where non-linear data trends correspond to year, biweekly, date-wise. The information is highly recommended to evaluate seasonal results and must have considering several seasons of past historical patterns. This model is completely automated which assists to get logical forecasting on unarranged data without manual support [14]. The Prophet model is accessible in Python and R language and used a similar standard code for the fitting of the model. It rapidly detects the changes in the linear, exponential, or logistic growth patterns by selecting switch points from the analyzed data. **Figure 3** corresponds to the components plot of the Prophet Model, which gives information about the model that is fitted.

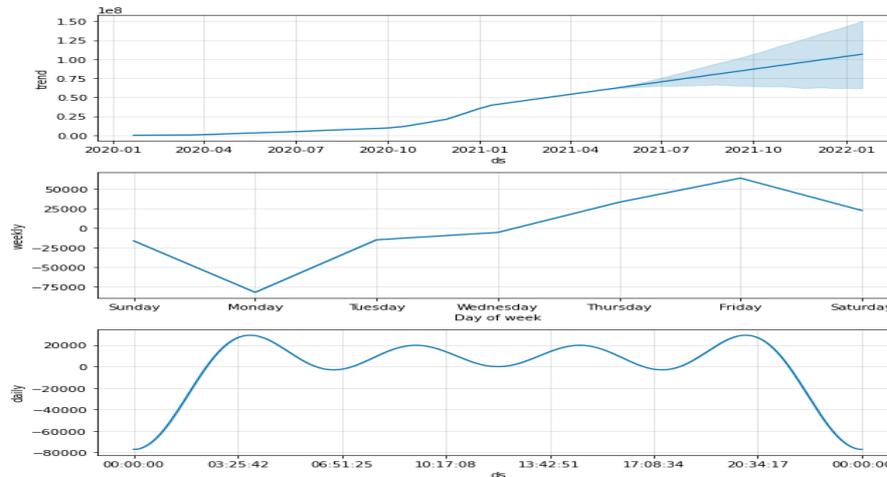


Figure 3: Prophet Component plot of the fitted model

4. Proposed Methodology

In the proposed study, the framework is used to extract and collect COVID-19 samples related data from multiple sources. The data is based on time-series analysis, which is meant to be constantly changing of data with time. In this study, the prophet model is fit the existing data including confirmed, recovered, death cases on the mentioned days. The time-series-based predictive trend curve indicates that the COVID-19 cases will increase or decrease over time in the future 260 days.

5. Results and Discussion

In this study, the plotted graphs for the confirmed, death, and recovered cases from the chosen sample dataset and compare these circumstances relating to various countries. Moreover, the visual inferences and statistical validation measures are performed by employing the “T-Test” in SPSS (Statistical Package for Social Sciences). The overall analysis of confirmed and recovered cases has been shown in **Figure 4**. **Figure 5** depicts the graphs of the corresponding confirmed, recovered, and death cases concerning their observation dates.

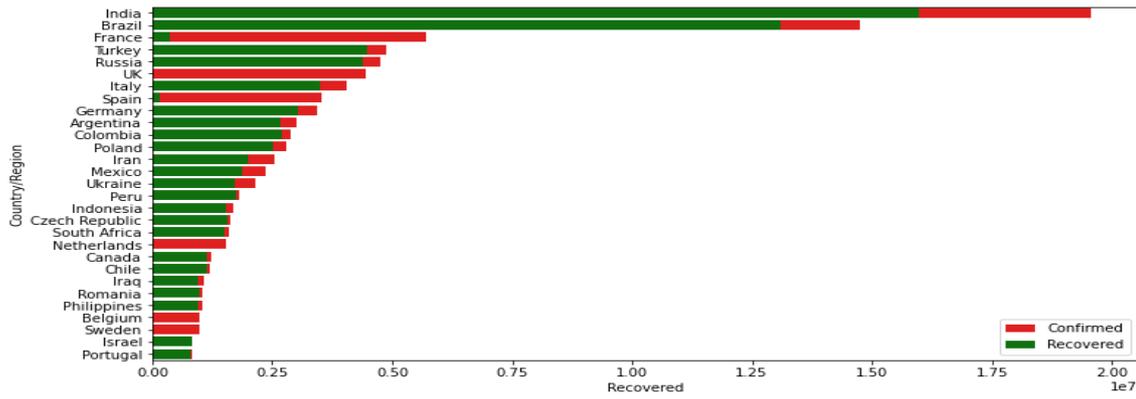
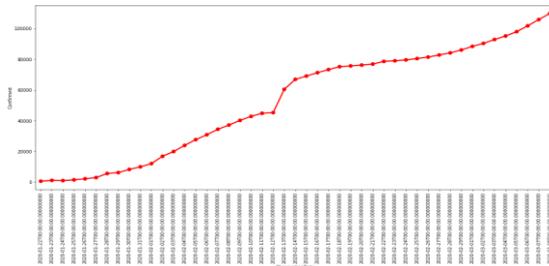
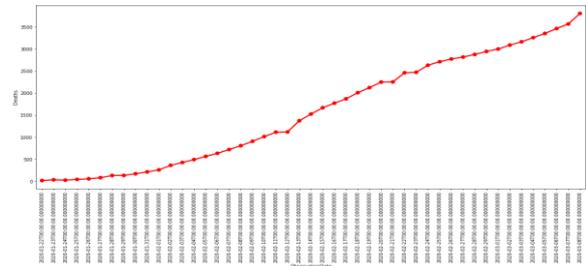


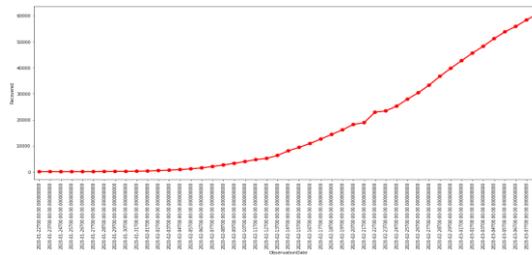
Figure 4: Confirmed and Recovered cases analysis based on different countries



(a) Confirmed Cases

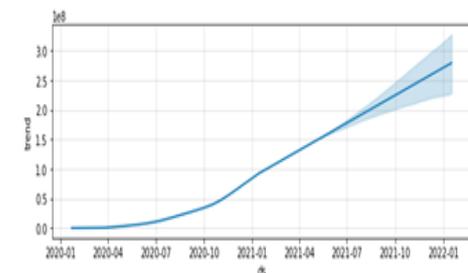


(b) Deaths Cases

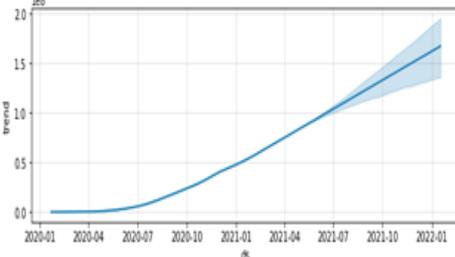


(c) Recovered Cases

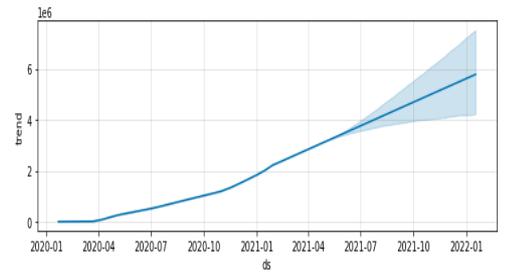
Figure 5: Graphical analysis of confirmed, deaths, and recovered cases from the existing data sample according to its observation dates



(a) Component plot for Confirmed Cases



(b) Component plot for Recovered Cases



(c) Component plot for Death Cases

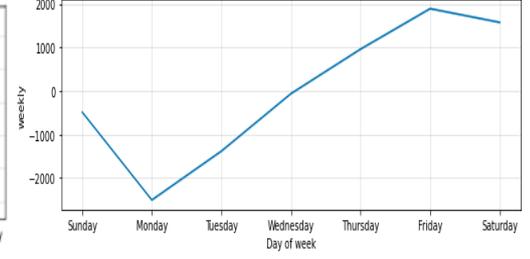
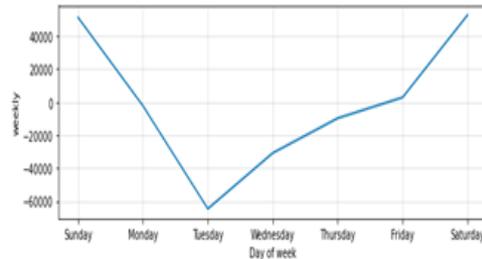
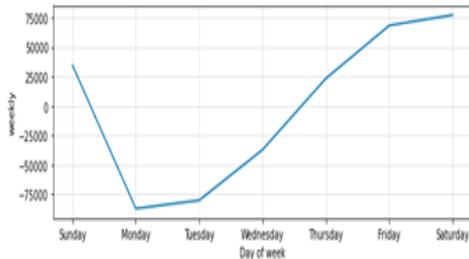


Figure 6: The component plot of Prophet Model for Confirmed, Recovered, and death cases

Figure 6 shows the curve of confirmed, recovered, and death cases from January 2020 to January 2022 in which it performs the time-series analysis for the next 260 days in various countries. Here x-axis denotes the dates and the y-axis denotes the cases where one unit represents 5, 00, 000 cases.

Table 1: T-Test applied for one-sample statistics

	N	Mean	Standard Deviation	Standard Error Mean
Confirmed_mean	187	91161.3659	182333.50909	13333.54672
Deaths_mean	187	1956.3371	4385.79371	320.72100
Recovered_mean	187	72819.7066	159262.98260	11646.46274

Table 2: One-Sample Test analysis

	T	DF	Sig. (2-tailed)	Mean-Difference	Confidence Interval (95%) with Difference
Confirmed_mean	6.837	186	.000	91161.36588	64856.9434
Deaths_mean	6.100	186	.000	1956.33715	1323.6187
Recovered_mean	6.253	186	.000	72819.70656	49843.5635

Table 1 and 2 shows the statistical analysis for one sample validation purpose by applying T-test approach. It performed on the mean value of confirmed, recovered, and death cases from the total sample. It achieves better validation results with a 2-tailed significant value ($P < 0.001$) so that the correlation is significant at a chosen confidence interval value.

6. Conclusion

In the situation of COVID-19, Predictive analysis strategies are followed to reduce the spreading rate of the pandemic. The prophet model is the most profitable way to forecast the future in a very efficient and accurate manner. This model framework is designed to recognize the infectious points from which the trend is highly variable and tackle outliers entirely. The statistical analysis will be of great significance to the official authorities, health departments, and medical organizations to produce drugs more quickly. This research provides a straightforward way to track the COVID-19 cases for forthcoming days at the global level. The paper describes the overall impact of lockdown extensions, social distancing, etc. to flatten the curve. One disadvantage of this approach, if the data have bi-weekly or quarterly time-series, then the framework will be inflexible to predict the future. To conquer this problem, all the suitable parameters must be configured manually.

References

1. WORLDOMETER (2020) COVID-19 coronavirus pandemic (2020). In: WHO. <https://www.worldometers.info/coronavirus/>. Accessed 5 Jan 2021
2. MERS-CoV (2020) WHO (Middle East respiratory syndrome coronavirus). <https://www.who.int/health-topics/middle-east-respiratory-syndrome-coronavirus-mers>. Accessed 9 Oct 2021
3. SARS-CoV (2020) WHO (Severe Acute Respiratory Syndrome). https://www.who.int/health-topics/severe-acute-respiratory-syndrome#tab=tab_1. Accessed 9 Oct 2021
4. CDC (2020) Coronavirus (Human Coronavirus Types). <https://www.cdc.gov/coronavirus/types.html>. Accessed 9 Oct 2021
5. WHO (2020) WHO Coronavirus Disease (COVID-19) Pandemic (2020). In: WHO. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed 30 Jan 2021
6. WHO (2020) WHO Coronavirus disease (COVID-19) dashboard (2020). In: WHO. <https://covid19.who.int/>. Accessed 29 Jan 2021
7. Anastassopoulou C, Russo L, Tsakris A, Siettos C (2020) Data-based analysis,

- modeling and forecasting of the COVID-19 outbreak. *PLoS One* 15:1–21. <https://doi.org/10.1371/journal.pone.0230405>
8. Alsaedy AAR, Chong EKP (2020) Detecting Regions At Risk for Spreading COVID-19 Using Existing Cellular Wireless Network Functionalities. *IEEE Open J Eng Med Biol* 1:187–189. <https://doi.org/10.1109/ojemb.2020.3002447>
 9. Jamshidi M, Lalbakhsh A, Talla J, et al (2020) Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment. *IEEE Access* 8:109581–109595. <https://doi.org/10.1109/ACCESS.2020.3001973>
 10. AL-Rousan N, AL-Najjar H (2020) Data analysis of coronavirus COVID-19 epidemic in South Korea based on recovered and death cases. *J Med Virol* 92:1603–1608. <https://doi.org/10.1002/jmv.25850>
 11. Lutz CS, Huynh MP, Schroeder M, et al (2019) Applying infectious disease forecasting to public health: A path forward using influenza forecasting examples. *BMC Public Health* 19:1–12. <https://doi.org/10.1186/s12889-019-7966-8>
 12. Guo C, Ge Q, Jiang H, et al (2020) Maximum Power Demand Prediction Using Fbprophet with Adaptive Kalman Filtering. *IEEE Access* 8:19236–19247. <https://doi.org/10.1109/ACCESS.2020.2968101>
 13. Chae S, Kwon S, Lee D (2018) Predicting infectious disease using deep learning and big data. *Int J Environ Res Public Health* 15: <https://doi.org/10.3390/ijerph15081596>
 14. Facebook Research (2020) Prophet: forecasting at scale. <https://research.fb.com/blog/2017/02/prophetforecasting-%0Aat-scale>. Accessed 9 Oct 2021