IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

Predicting Covid-19 Case Trend using Time Series Analysis Methods

Prof. Dinesh B. Satre, Mrunal R. Yemale, Namrata A. Kassa,
Harshada N. Mankeshwarker, Rutuja N. Lokhande
Department of Computer Engineering
Marathwada Mitramandal's Institute of Technology, Pune, India

Abstract: *The COVID-19 pandemic has created havoc all over the world. Millions of lives have been gone and thousands arevulnerable. It has also affected the world economy due tolockdown. So, there is a need to develop a time-series forecasting model for predicting future cases so that necessary precautions can be taken. The aim is to help incoping up with the situation without affecting lifestyle anyfurther. Therefore, an accurate prediction of the future spread of COVID-19 becomes crucial in such a situation. In this comparative study, two different time-series analysis models, namely the ARIMA model, the Prophet model, which are machine learning model are investigated to determine which has the best performance when predicting the future case trends of COVID-19 in India. The project highlight is to predict the spread of COVID-19 so that countries can be better prepared and aware when controlling the spread.*

Keywords: COVID-19, Time-Series Forecasting, ARIMA , Prophet, machine learning

I. INTRODUCTION

COVID19, which is the significant wellspring of sicknesses going from gentle colds to more intense infections, for example, MERS-CoV and SARS-CoV, as indicated by the World Health Organization (WHO 2020). Another Covid (nCoV) variety has not been found in people before. Coronavirus impacts affect other segments too, for example,the old and the hidden medical conditions.The ongoing COVID19 pandemic has created havoc all over the world. Millions of lives have been gone and thousands are vulnerable. It has also affected the world economy due to lockdown.

Because of the quickest spread of the new COVID-19, the combined number of affirmed cases and the everyday number of new cases are as yet expanding in the nations. Hence, an investigation of the momentum total and the new number of instances of Coronavirus has fundamental examination suggestions for anticipating its pervasiveness patterns.

Various models have been utilized to anticipate Coronavirus predominance and death rate in research studies. For instance,multiple linear regression, Artificial Neural Network, multilayer perceptron grey prediction model, simulation model, Holt model, LSTM model, and support vector regression. However, the spread of epidemic disease is random and will be affected by many factors. A large numberof studies all show that the effect is not best achieved if only a single prediction tool is utilized to predict trends. In addition, the above statistical model can predict the development trend of the epidemic in the medium and long term.

The Automatic Regressive Integrated Moving Average (ARIMA) model has some advantages in its simple structureand immediate applicability. The ARIMA model has been applied to the prediction and estimation of prevalent diseases,such as typhoid fever, tuberculosis, influenza and COVID-

Since ARIMA methods do not contain much mathematicsor statistics, but also are capable of correlating regulation with short-term changing trends in the time series. So, the model is more suitable for predicting the short-term epidemic diseases.

The Prophet is an open source model that can handle time- series data with the advantages of taking strong seasonal effects, missing data, outliers, and changes in trends. And it is currently useful for predictive analysis of COVID-19. What's more, the SARIMA and Prophet models can be usedto capture some periodic or seasonal changes, further find thenonlinear fluctuations of data, and improve the accuracy of prediction results.

Subsequently, it is of incredible down to earth importance toanticipate the day to day new cases and combined affirmed instances of Coronavirus from one side of the planet to the other. This study lays out the ARIMA and Prophet models to

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

foresee the day to day new cases and combined affirmed instances of Coronavirus in the USA, Brazil, India and assess the expectation exactness of the model to give a further reference to the forecast and early admonition of irresistible sicknesses

II. LITERATURE REVIEW

In this paper, an experimental study was conducted by Aayush Jain, Tanay Sukhdeve, Himanshu Gadia, Dr. Satya Prakash Sahu, Satya Verma for forecasting of COVID-19 pandemic spread pattern. They did comparison of the difference between forecasted and original value using RMSE and MAE. It was found that the ARIMA model with parameters (p,d,q) = (8,2,1) worked best among all with RMSE of 2773.27.

In this paper, a comparative study, they investigate the performance of five different time-series analysis methods (including a naïve baseline model) when applied to predict COVID-19 cases in six different countries. And according to this they concluded that LSTM model performed best amongthe all.

In this paper, Yanding Wang, Zehui Yan, Ding Wang, Meitao Yang, Zhiqiang Li, Xinran Gong, Di Wu, Lingling Zhai, Wenyi Zhang and Yong Wang does comparative study for forecasting the future result of COVID-19 cases using ARIMA and Prophet models. Further, the prediction of the Prophet model showed sufficient accuracy in the daily COVID-19 new cases of the USA. The ARIMA model is suitable for predicting Brazil and India, which can help take Precautions and policy formulation for this epidemic in othercountries.

Hadeel I. Mustafa et al. [8] did forecasting based on a dataset published by the Iran health ministry regarding the COVID19 breakout. ARIMA (2,1,5) model gives the best prediction results with the performance evaluation being done by MSE and MAE.

III. PROPOSED SYSTEM

3.1 System Architecture

The major components of our application are shown inFig.1.

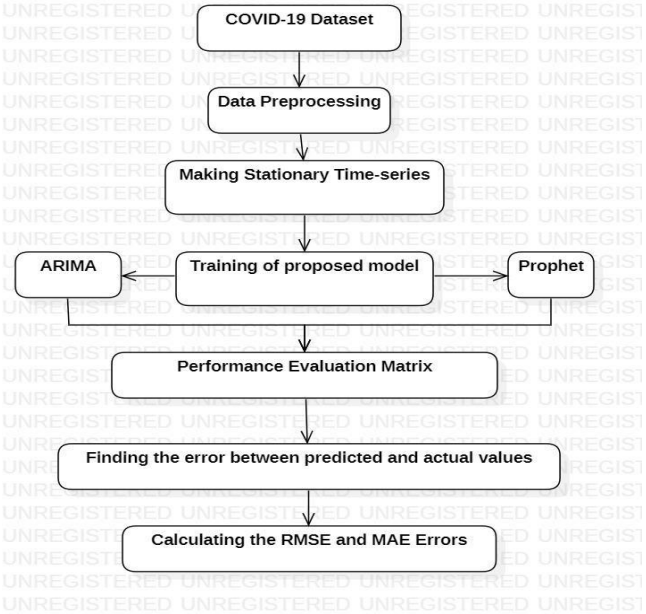

Fig 1. System Architecture

IV. METHODOLOGY

4.1 Data Collection

The data used in this project is collected from kaggle.com i.e COVID-19 cases worldwide-cases.csv dataset. It provides accurate statistics of case reports in the US and many other countries, offering a wide range of choices. Thedataset also contains a time-series summary of COVID-19 cases, which is very suitable for our project research.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

4.2 Data Pre-Processing

Due to the amount of cases, we decide to use daily new casesas the input data (or the training data) for all time-series analysis models. We notice that there are negative values in the dataset with some minor outliers. To resolve the problem,we can simply remove the negative values from the data andreplace it with the data on the previous day. From a statistical perspective, it eliminates the errors that will confuse the models and it provides a smoother and reasonable curve for time-series models. The purpose is to test the performance ofmodels in predicting the trend of COVID-19, but not the totalcases, thus removing negative values in this way is appropriate in our project.

4.3 Making Data Stationary

We are using rolling statistics method for making datastationary.
Rolling Statistics:- Plot the rolling average. Rolling average is one of the most common techniques in time-series analysis. It is used to smooth the data and distinguish the noise within the time-series sequence, which prevents any analytical models from performing poorly. Rolling average usually means replacing the value with the mean value of its n neighbours. This helps in making data stationary which is useful to give accurate time series forecast. It is stationary ifthey remain constant with respect to time.

4.4 Experiment Models

Autoregressive Integrated Moving Average (ARIMA)Model:
It can be considered as the generalized version ofAutoregressive Moving Average (ARMA) that is built by combining the Autoregressive (AR) process and MovingAverage (MA) process and builds a compound model of the time series. As indicated by the acronym, ARIMA(p, d, q) has the following as key elements of the model:

AR (Autoregression) : It works on the principle that variable of interest is regressed on its own laggedobservations (p).
I (Integrated) : To convert the time-series into stationary form by removing the trend. It is done by calculating the difference of data at different points (d).
MA (Moving Average) : A method based on the relationship between the actual data and the error values when a model based on moving average is applied to the number of lagged values (q).
The equation for ARIMA is given by,

$$Yt=\alpha+\beta 1Yt-1+\beta 2Yt-2+\cdots+\beta pYt-p+\emptyset 1\epsilon t-1+\emptyset 2\epsilon t-2+\cdots+\emptyset q\epsilon t-q$$

Prophet

Prophet is a forecasting procedure released by Facebook's Core Data Science team in 2017. It is an open source software and is available in both Python and R programming language.Prophet can perform at its best when the data have strong seasonal effects, and it handles missing data or data shifts in the series very well. Prophet is accurate and fast, and it provides multiple user-friendly parameters to tune the model,adapting to specific domains to improve the performance.
To prevent possible negative predicted values, some measures need to be performed. We first take the natural logarithm of data as input (training data) to feed into the model, and then take natural exponential of the output (predicted values) as the result. Negative infinity is treated as NaN to avoid mathematical problems. The formulation ofProphet can be described as follows:

$$y(t) = g(t) + s(t) + h(t) + t\ .$$

g(t) is for demonstrating non periodic alterations in period series, **s(t)** are the changes that are done periodically like weekly, yearly, and seasonally, **h(t)** is the effect of holidaysthat the user gives with irregular schedules.

Evaluation Metrics

Evaluation metrics used in this project are Mean SquaredError (MSE) and Mean Absolute Error (MAE).

Mean Squared Error (MSE)

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

Mean Absolute Error(MAE):-

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right|$$

where ^yi is the predicted value and yi is the actual value. These two metrics can represent the accuracy of models well and can be used to evaluate their performance. For each model, we will calculate these two metrics for the values in the results and compare with each other, determining the model with best accuracy. Other evaluation metrics can also be used in this scenario, but to the simplicity of evaluation wewill only use MSE and MAE.

UML Diagrams

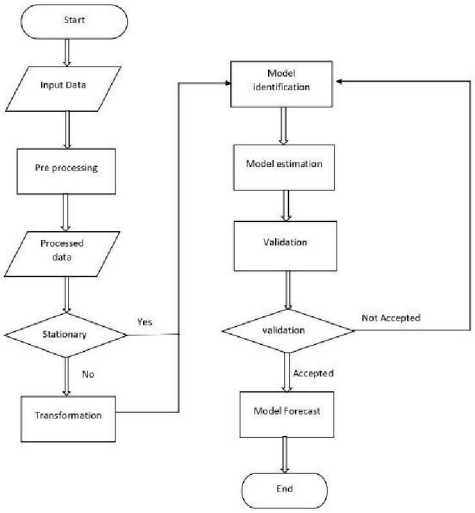

Fig. 2. Activity Diagram

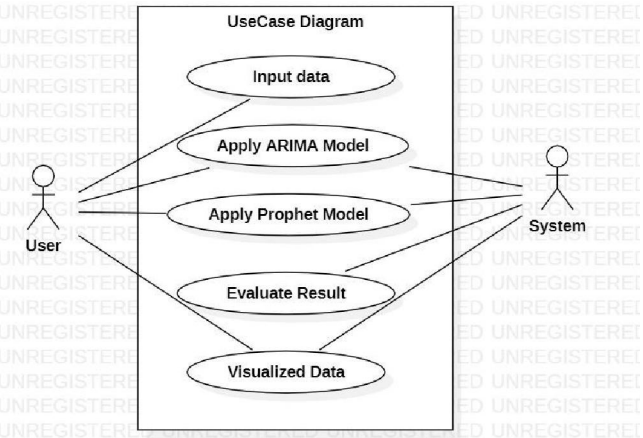

Fig 3. Use-Case Diagram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

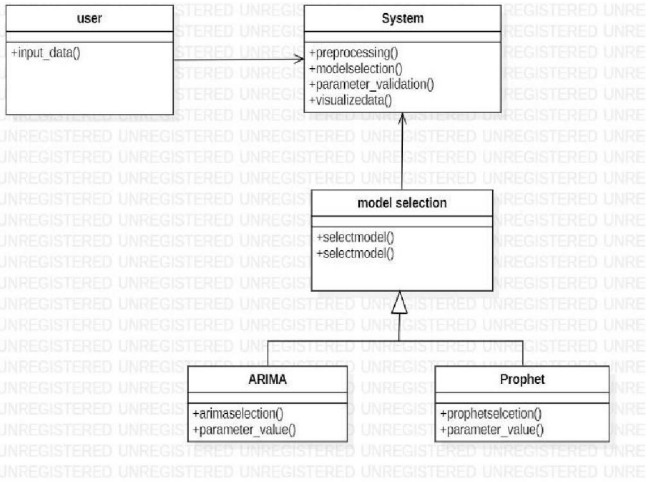

Fig 4. Class Diagram

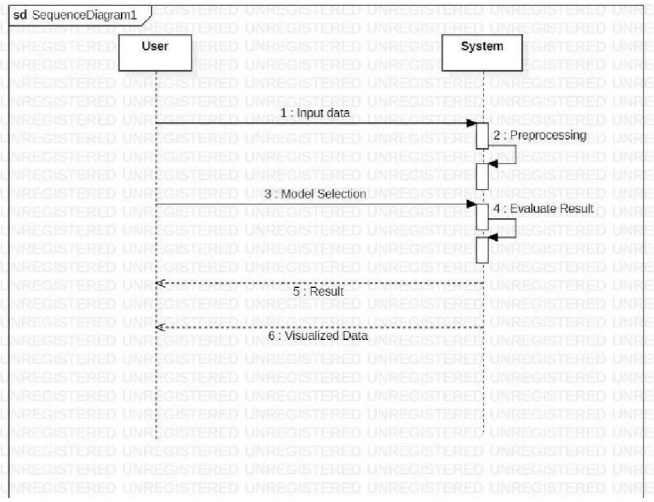

Fig 5. Sequence Diagram

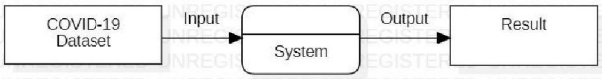

Fig 6. Data Flow Diagram Level 0

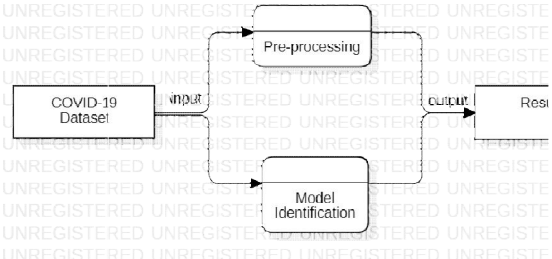

Fig 7. Data Flow Diagram Level 1

.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

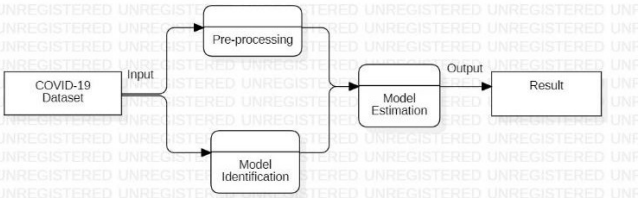

Fig 8. Data Flow Diagram Level 2

V. CONCLUSION

In this comparative study, we explore the presentation of twodifferent time-series examination techniques when applied to foresee Coronavirus cases in four distinct nations. We first gather the COVID19 case information from world data fromhttps://kaggle.com/Covid-19 and select India country with the most aggregate affirmed cases. Then we acquire the day to day new cases for every one of the nations and smooth the time-series information with pre-processing techniques. After normalizing the information, we apply two distinct techniques, to be specific an essential gauge model, the ARIMA model, the Prophet model, to fit and foresee the future case number and pattern in these nations. By contrasting the two measurements of the expectation results we can anticipate which model will be more compelling to conjecture the future for Coronavirus. Apart from that, a comparison of the difference between forecasted and original value was done. The linearity of COVID patterns can be easily captured using an efficient linear model.

ACKNOWLEDGMENT

In the accomplishment of this project successfully, manypeople have extended a helping hand, we would like to show our deep appreciation for them and we are utilizing this timeto thank all the people who have been concerned with the project. I would like to thank our Principal, Dr. R. V. Bortake, and our HOD, Prof. Subhash G. Rathod, for providing us with the golden opportunity to work on this project.

Their suggestions and instructions have served as the major contribution towards the completion of this project. We would also like to thank our guide, Prof. D. B. Satre, for his valuable support and guidance throughout the completion of our project. We would also like to thank our classmates who have helped us with their valuable suggestions and support which has been very helpful in the completion of this project

REFERENCES

[1] 10101998 Coronavirus Update (Live). Cases and 501644 deaths from COVID19 virus pandemic - worldometer. Available at: https://www.worldometers. info/coronavirus/. [Accessed 28 June 2020].

[2] lok Kumar Sahai , Namita Rath , Vishal Sood , Manvendra Pratap Singh. ARIMA modelling & forecasting of COVID-19 in top five affected countries.

[3] Sulasikin, Y. Nugraha, J. Kanggrawan and A. L. Suherman, "Forecasting for a data-driven policy using time series methods in handling COVID-19 pandemic in Jakarta," 2020 IEEE International Smart Cities Conference (ISC2), Piscataway, NJ, USA, 2020, pp. 1-6, doi: 10.1109/ISC251055.2020.9239066.

[4] Andres Hernandez-Matamoros, Hamido Fujita, Toshitaka Hayashi. Forecasting of COVID19 per regions using ARIMA models and polynomial functions

[5] S. Siami-Namini, N. Tavakoli and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, 2018, pp. 1394-1401, doi: 10.1109/ICMLA.2018.00227.

[6] V. Kotu and B. Deshpande, “Chapter 12 - Time Series Forecasting,” in Data Science (Second Edition), V. Kotu and B. Deshpande, Eds. Morgan Kaufmann, Jan. 2019, pp. 395–445. ISBN 978-0-12-814761-0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978 0128147610000125 [Page 8.]

[7] “Prophet.” [Online]. Available: http://facebook.github.io/prophet/ [Pages 9 and 30.]

[8] I. Yenidoğan, A. Çayir, O. Kozan, T. Dağ, and Arslan, “Bitcoin Forecasting Using ARIMA and PROPHET,” in 2018 3rd International References | 57 Conference on Computer Science and Engineering (UBMK), Sep. 2018. doi: 10.1109/UBMK.2018.8566476 pp. 621–624. [Page 10.]

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

[9] G. A. Papacharalampous and H. Tyralis, "Evaluation of random forests and Prophet for daily streamflow forecasting," in Advances in Geosciences, vol. 45. Copernicus GmbH, Aug. 2018. doi: 10.5194/adgeo-45- 201-2018 pp. 201–208, iSSN: 1680-7340. [Online]. Available: https://adgeo.copernicus.org/articles/45/201/2018/ [Page 10.]
[10] C. B. Aditya Satrio, W. Darmawan, B. U. Nadia, and N. Hanafiah, "Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET," Procedia Computer Science, vol. 179, pp. 524–532, Jan. 2021. doi: 10.1016/j.procs.2021.01.036. [Online].Available: https://www.sciencedirect.com/science/article/pii/ S1877050921000417 [Page 14.]