COVID-19 Prediction Using Time Series Forecasting Models

Himanshi
dept. of CSE
IIT Jodhpur
India
himanshi.2@iitj.ac.in

Shubham Singh
dept. of CSE
IIT Jodhpur
India
singh.160@iitj.ac.in

***Abstract*—COVID-19 is an infectious disease caused by the SARS-CoV-2 virus. The first case of COVID-19 emerged from China on Dec. 2019. Millions of people have been affected by this virus and lakhs of people lost their lives, their jobs due to this ongoing Novel COVID-19 Pandemic. It is of absolute importance to identify the future infected cases and the virus spread rate for prior preparation in the healthcare services to avoid deaths. It is the most challenging real-time problem for forecasting the spread of COVID-19. We will be using the COVID-19 dataset that is available on Kaggle and we will perform predictions using ARIMA and Prophet time series forecasting models.**

I. Introduction

The corona virus infected millions of people globally due to its high community impact and extensive spreading. There is a need to recognise future infected cases. We perform Modelling and forecasting on the coronavirus's spreading behaviour so that the health systems manage the forthcoming number of cases. Correct forecasting of the pandemic is a concern because it influences the government policies, health system, containment customs, and social life. Concerning this, we test the capability of the ARIMA and FBProphet forecasting models. The models are used and followed due to their absolute forecasting ability.

II. Modeling Dataset

The data is collected from [1], "Novel Corona Virus 2019 Dataset". This dataset contains the number of COVID-19 cases between 22 Jan, 2020 to 29 May, 2021 and was last updated on 24 June, 2021. We use the day-wise and country-wise collective cases of COVID-19, as shown in Fig. 1, Fig. 2, and Fig. 3.

III. Time series forecasting models

Our report intends to present an evaluative study of predicting COVID-19 cases and anticipate the influence of the virus on the infected countries. We examined the execution of two models Arima and FBProphet, and calculated the root mean square error (RMSE). We also presented the forecasting results for COVID-19 concerning both models.

Country/Region	Confirmed	Recovered	Deaths
US	33251939.0	0.0	594306.0
India	27894800.0	25454320.0	325972.0
Brazil	16471600.0	14496224.0	461057.0
France	5719877.0	390878.0	109518.0
Turkey	5235978.0	5094279.0	47271.0
...	...	...	...
Vanuatu	4.0	3.0	1.0
Marshall Islands	4.0	4.0	0.0
Samoa	3.0	3.0	0.0
Kiribati	2.0	0.0	0.0
Micronesia	1.0	1.0	0.0

195 rows × 3 columns

Fig. 1. Country-wise grouped dataset.

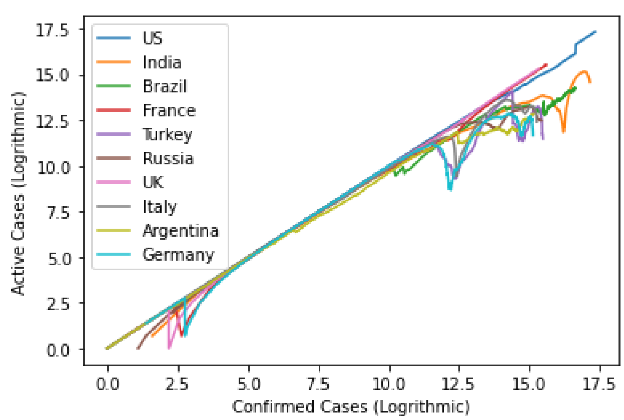

Fig. 2. Country-wise grouped active and confirm cases plot.

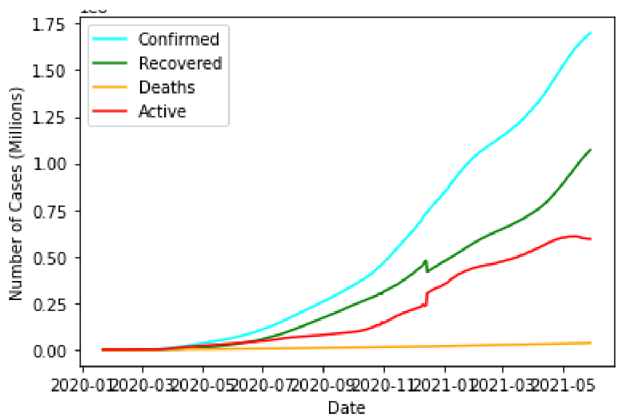

Fig. 3. Date-wise grouped Plot.

A. Arima Model

Arima stands for an autoregressive integrated moving average. The ARIMA model converts a time series into a series without any trend or seasonality using differencing, which implies that the mean and variance for the series is constant[2]. The equation of the Arima model is:

$$z_t = \alpha + \sum_{i=1}^{p} \phi_i z_{t-i} + w_t + \sum_{j=1}^{q} \theta_j w_{t-j} \tag{1}$$

here, ϕ_i are lag coefficients which are estimated by the model, z_{t-i} are the lags (past values), w_t is the white noise, w_{t-j} are error terms of the model for the respective lags i.e, z_{t-j}, and α is defined as follows [3].

$$\alpha = \left(1 - \sum_{i=1}^{p} \phi_i\right) \mu \tag{2}$$

here, μ is mean of the process [3].

B. FBProphet

FBProphet (Facebook Prophet) uses linear and non-linear methods as the segments with time. The prophet was developed and released as open-source software by the data science team of Facebook. The model overlooks the temporal dependence of provided data, and the training part is framed as a curve-fitting operation[3]. The equation of the Prophet model is:

$$z_t = T_t + S_t + H_t + \epsilon_t \tag{3}$$

here, T_t is trend, S_t is seasonality, H_t is holiday, and ϵ_t is error term [3].

C. Forecasting Framework

- First we divide our dataset into training and testing. Training and testing splitted into the ratio 95:5. we have also tried different splits but couldn't get better accuracy then this one.
- Then we try to predict covid-19 cases using ARIMA Model using autoArima. The loss function we have used is root mean squared error (RMSE). Then we try to predict the covid-19 confirmed cases as well as Death Cases.
- We also have predicted the confirmed cases using FBProphet and measured the loss using RMSE.
- By doing so, we found out that rmse score for prophet is less then the arima model. Meaning performance of Prophet Model is better then ARIMA Model.

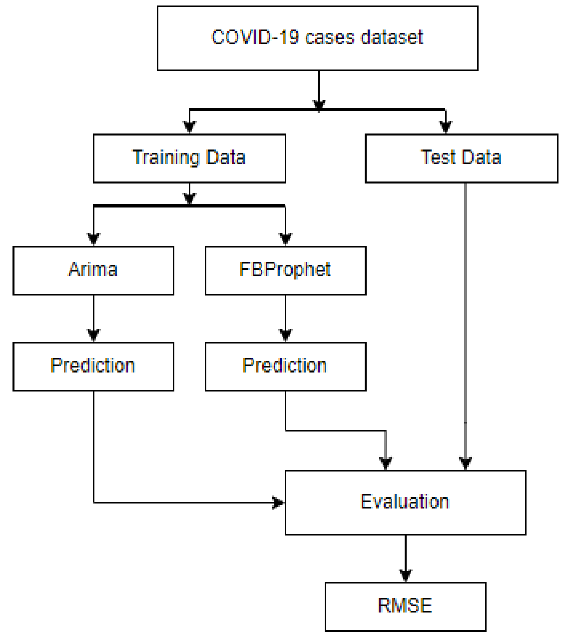

Fig. 4. Forecasting framework.

IV. Results

The framework is implemented in Python 3.8, and we used the ARIMA and the Prophet models to forecast predictions. We completed our experiments in the Intel Core i3 processor clocked at 2.00 GHz, 8 GB RAM. This section includes the forecasting accuracy of the models for confirmed and death cases.

A. Forecasting using ARIMA model

Using the Arima model, we analysed the confirmed cases and death cases trend concerning the training data, test data and the cases predicted by the model. Fig. 5 and 6 shows the plot of the confirmed cases and death cases, respectively. From where we can recognise that the predictions are very close to the actual data, and the model fits well.

B. Forecasting using Prophet model

Using the Prophet model, we analysed the confirmed cases and the confirmed cases weekly trend concerning the training data, test data and the cases predicted by the model. Fig. 7 and 8 shows the plot of the confirmed cases and weekly cases, respectively. From where we can recognise that the predictions are approaching the actual data.

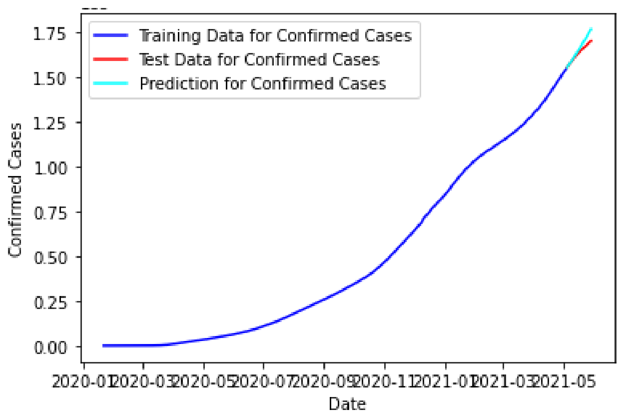

Fig. 5. Prediction of confirmed cases using ARIMA model.

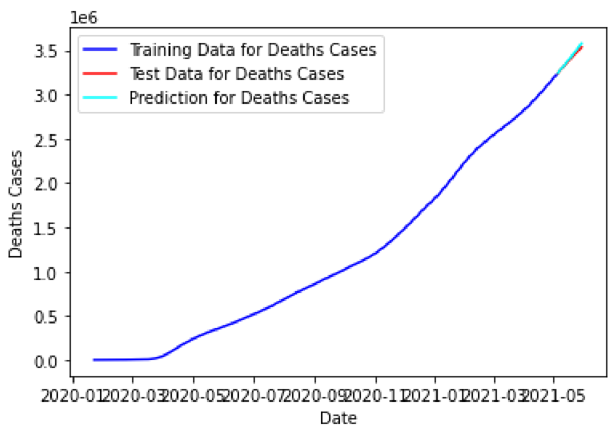

Fig. 6. Prediction of deaths cases using ARIMA Model.

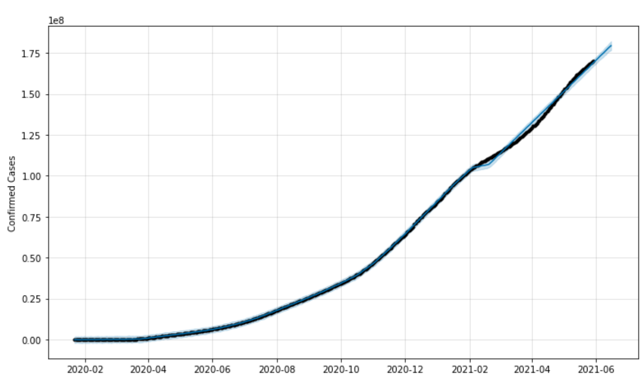

Fig. 7. Prediction of confirmed cases using Prophet model.

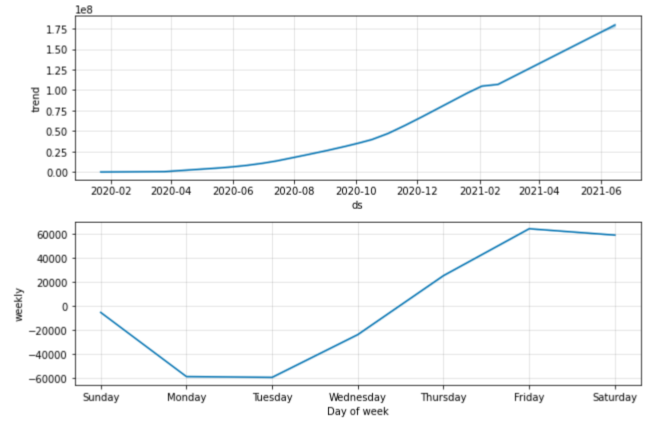

Fig. 8. Prediction of confirmed cases component using Prophet model.

C. Performance Measures

We calculated Root Mean Square Error(RMSE) using the equation 4 to measure the performance of the models. Fig. 9 shows the RMSE values and indicates that the FBProphet model is more accurate than the Arima model on the chosen dataset.

$$RMSE = \sqrt{\frac{1}{N}\Sigma_{i=1}^{n}(e_i)^2} \quad (4)$$

Model Name	RMSE
Facebook's Prophet Model	1.025041e+06
ARIMA Model	3.160138e+06

Fig. 9. Performance results.

References

[1] https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset/code

[2] https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06

[3] N. Kumar and S. Susan, "COVID-19 Pandemic Prediction using Time Series Forecasting Models," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225319.