iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VIII **Month of publication:** August 2020

DOI: https://doi.org/10.22214/ijraset.2020.30784

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

Analysis and Forecasting COVID-19 in India

J S Anandatirtha[1], B Harini Shree[2], M Janani[3]
[1, 2, 3] *Students, Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Sriperumbudur, Kanchipuram, Tamilnadu, India.*

Abstract*: The world is dealing with an imperceptible, fatal enemy and enforcing arduous efforts to cope up with menace due to COVID-19. This worldwide health crisis posed a unique challenge in tracking the patients, building dedicated hospitals, purchasing PPE kits, ventilators, and medicines. In this research, the study on the actual situation in India was conducted with the available datasets. In addition, how this pandemic condition would increase in the forthcoming days is predicted through a widely popular approach called Forecasting. Forecasting is a way to predict what is going to occur in the future by taking into consideration the events in the previous and current models. The most prominent forecasting algorithm is deployed on the COVID-19 dataset to forecast how the cases will vigorously increase till November 2020. Moreover, these algorithms are evaluated using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE).

Keywords: Time series, Corona virus, Forecasting, Prophet, WHO.

I. INTRODUCTION

COVID -19 (shortened of Corona Virus 2019) is an infectious disease that affects humans and causes ailment ranging from the common cold, breathlessness, muscle pains to Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS-COV), in some cases even death [8,9]. COVID-19 had a tremendous impact on the society that affect even the most established medical facilities to come to a standstill. In December 2019, multiple pneumonia cases had reported, and eventually, the Chinese Government had spotted the first corona case in Wuhan, Hubei Province [7]. The virus spread through droplets from the infected patients and touching the surface contaminated with the virus [11]. According to the World Health Organization (WHO), a total of 11.8 million people was affected due to COVID-19. The total recovered from the cases was about 6.42 Million people as of 2/7/2020. The world had encountered various infectious diseases such as Influenza, Bubonic Plague, Smallpox, Measles, Yellow fever, Typhus, Influenza A virus subtype H2N2, and Tuberculosis impacted people's life. COVID-19 as the pandemic situation where the spread of the diseases is throughout the globe and to be handled carefully. Due to the high dense population and contamination of infectious viruses, countries like South Africa, Saudi Arabia, India, Jordan, China, Australia, the United States of America, Israel, France, Spain, etc enforced series of Lockdown. There is no medication or vaccination to control the spread of the infectious virus [10]. The major countries like India, the US, China, Russia, and England are involved in the development of vaccines to control the spread. In this era, the modern sector target on the Time series Forecasting algorithms considering the wider application in Weather Forecasting, Earthquake prediction [5], sales forecasting [13], Technology forecasting [6], conflict situations [14], and much more. Time series analysis is a way to investigate the sequence of time intervals to extract some meaningful insights. Forecasting is an approach to predict the future based on the quantitative value, including historical data and knowledge. It depends on a mathematical model making it efficient and persistent by applying a generous amount of data rather than an expert's speculation. It provides valuable information regarding any business trends, helps to invest in stocks, able to forecast the behaviour of diseases, predict the weather, etc by examining historical data.

One of the essential properties of Time Series is stationary, where the statistical properties remain persistent for some time. The data can observe to investigate the trends, cycles that further help to enhance the forecasting of an event. The trend can be observed when grow or decline in the time series. The constituent of the time series is Trend, Seasonality, Level, Noise. Depending on the frequencies, the time series will represent daily, weekly, monthly, yearly. Based on trend and seasonality, a time series can be distinguished through additive and multiplicative time series.

A. Additive time Series

Model = Base level + Trend + seasonality + error (1)

B. Multiplicative time Series

Model = Base level * Trend * seasonality * error (2)

©IJRASET: All Rights are Reserved

This research uses time series forecasting as it uses highly sophisticated and precise information about the correlation among the factors. For time series analysis and projection techniques, the historical data is significant for prediction models. Time series analysis provides a strong knowledge about the patterns in the data and growth rates of these trends and helps in achieving better results. Also, it is necessary to forecast accurately to prevent the critical situation and to save on time and cost.

II. LITERATURE SURVEY

In 2016, Kalid Yunus proposed a new methodology by administrating a revised ARIMA approach that can capture time correlation and the probability distribution of observed wind speed time-series data was demonstrated. This approach establishes frequency decomposition (by parting the data into high frequency and low-frequency components), shifting, limiting, transformation is used in the ARIMA modeling procedure [1]. In 2016, Yanchun Pan et-al identified the ARIMA model to forecast the number of epidemic diseases. Firstly, Autocorrelation and partial Auto Correlation were introduced to establish a stationary time series. Then, Least Squares is used to determine the parameters of the model. Finally, the data between January and August were fed into the model and obtained an accuracy of 92.1%. This model yields a comparatively better result than the model used by the CDC [2]. Yenidog et-al presented a paper to forecast the bitcoin with prophet and ARIMA models. Also compared with each other in terms of metrics. Time series analysis had a univariate characteristic and added with some additional variables to each model to improve the forecasting accuracy. These values are based on the different correlations between cryptocurrencies and real currencies. The model selection is carried out using three folding splitting technique. Here prophet outperforms ARIMA by 0.94 to 0.68 in R2 [3]. Fang et-al proposed a methodology for forecasting the financial markets by employing LSTM and prophet to predict the trend. Then, the prediction trend is combined with an inverse neural network for prediction. Overall, outcomes show that the method can achieve accurate forecasting with fewer errors [4].

III. ANALYSIS OF COVID-19 IN INDIA

According to Johns Hopkins Corona Virus Resource center, more than 187 countries were affected due to COVID -19. In India, the first case spotted in Kerala on January 2020. The confirmed cases in India gradually increased and reached 6, 21,000 as of 2/7/2020, while the death rate is nearing 18,000. Indian Government had allocated INR 3000 cr from the PM-Cares fund to resist the COVID by purchasing essential equipment. Moreover, the Government had allocated INR 100 cr for the development of vaccines.

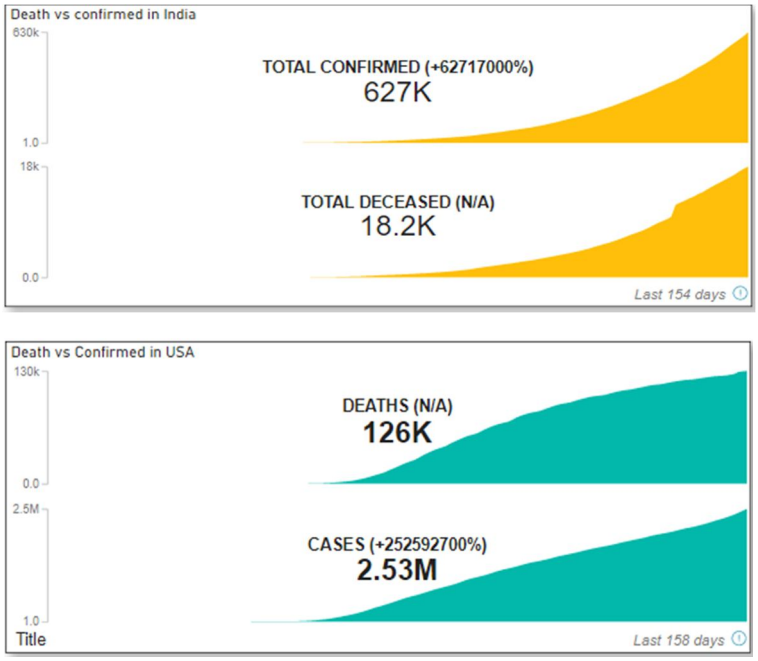


Figure 1. Comparison of Confirmed and Deceased cases between India and USA

Figure 2 illustrates the number of populations in India with the corresponding state. Overall, the total population in Uttar Pradesh is higher. Also, the total population in Maharashtra and Bihar is more or less equal.

©IJRASET: All Rights are Reserved

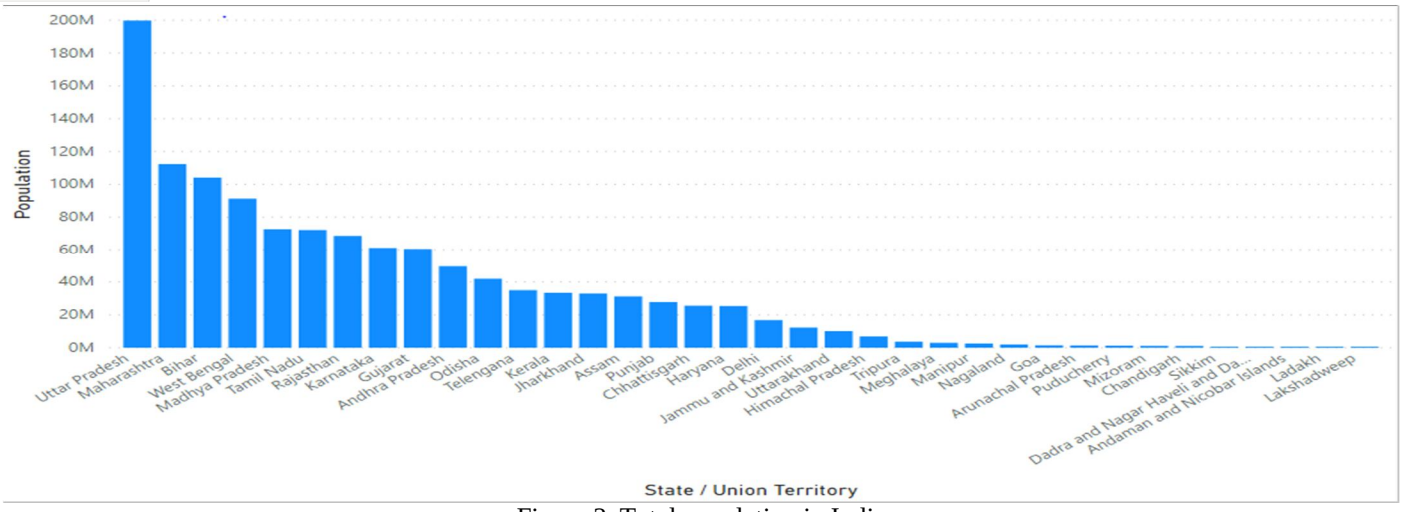


Figure 2. Total population in India

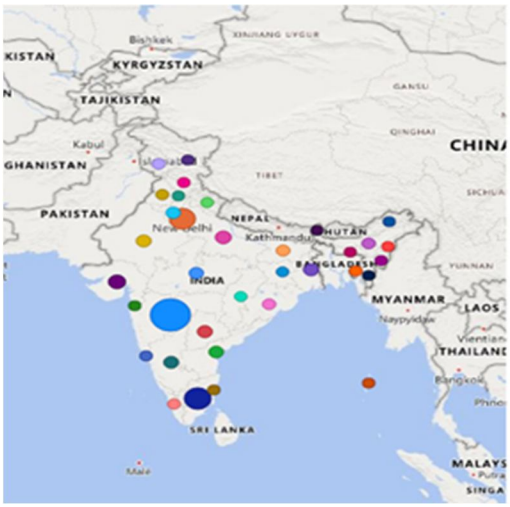


Figure 4. Severity in confirmed cases in India

According to the 2011 census, India had a total population of 1 Billion. Due to the enormous population, it is hard to control the spread. Figure 3 illustrates the states with higher cases, while Maharashtra has a higher rate. Next, Tamilnadu and Delhi have an equal similar to confirmed cases. Figure 4 explains the heat map of India based on the total number of cases. In India, the confirmed cases are gradually increasing due to various causes such as direct contact with infected peoples, gatherings, people arriving from abroad, or from containment areas.

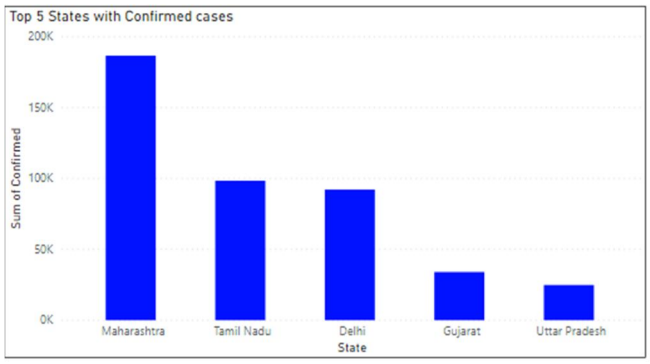


Figure 3. Top 5 states in confirmed cases

©IJRASET: All Rights are Reserved

While discussing recovery, India has a recovery rate of 61.18 % of the total cases, which is comparatively higher than in other countries. India has an active case of 2, 23,000 as of 2/7/2020. Maharashtra, Delhi, Tamilnadu, Gujarat, and Uttar Pradesh are the states with a higher recovery rate.

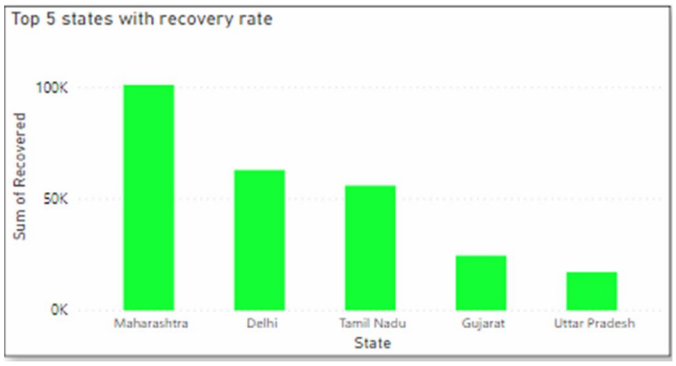


Figure 5. Top 5 states in Recovery rates

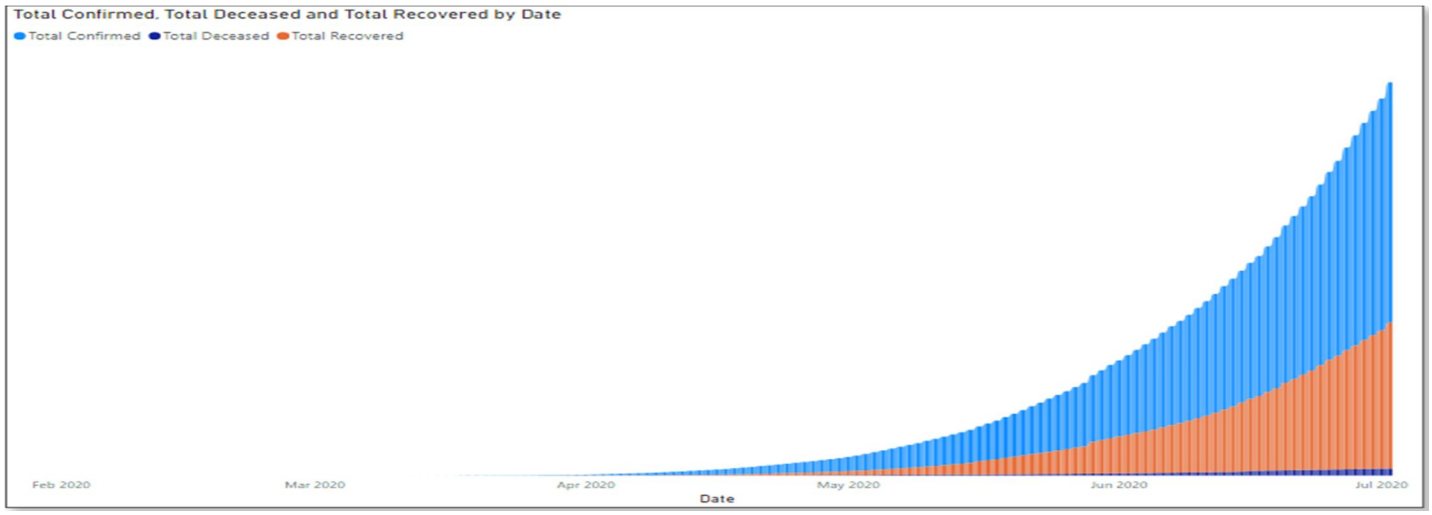


Figure 6. Total Confirmed, Deceased, Recovered cases in India

Considering death cases, the Government had reported 18,000 deaths as on the 7/2/2020. Oldage people are more vulnerable to coronavirus because they do not have a strong immune system to fight against infectious diseases. Cancer, Heart disease, lung disease, diabetes, lower respiratory problem, hypo-thyroid, Influenza and Pneumonia, Cerebrovascular diseases are the most common pre-existing underlying medical conditions in corona virus-related deaths.

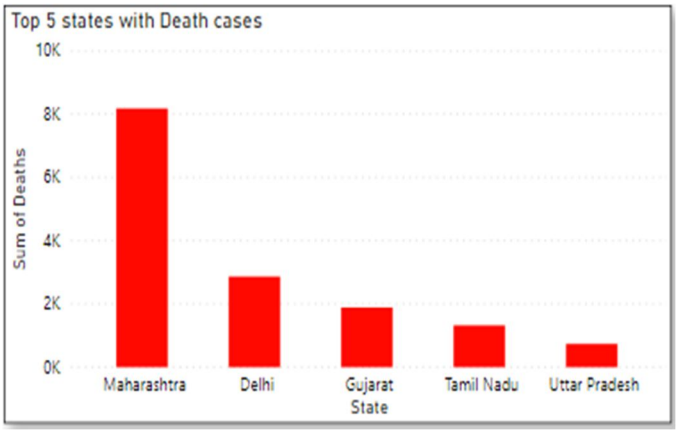


Figure 7. Top 5 states in death rates

©IJRASET: All Rights are Reserved

India follows some strategies to prevent the spread of diseases.

1) By closing down the Educational Institutions, IT parks, Industries, Hotels, Temples, Marriage Halls, Markets, Malls, and even salons to avoid mass gatherings.
2) By closing the state and national borders.
3) By screening and isolating the people traveling from abroad for 14 days.
4) By vigorously tracking the contact history of infected persons.
5) By imposing red, green, orange zones within districts.
6) Public transportations such as Metro, Bus, Trains, and Flights stopped operating.

IV. METHODOLOGY

A. Data Acquisition

In this research, the COVID-19 dataset obtained from the Kaggle website consists of 156 rows and three columns. The features of the dataset are Confirmed, Deceased, and Recovery day to day cases in India with the corresponding date. It contains daily reports of COVID-19 from January 2020 to June 2020. Also, the population dataset to analyze the population in India obtained at the time of the 2011 census.

B. Proposed Model

The Forecasting approach involves the number of stages fig 8 demonstrates the overall structure for time series forecasting.

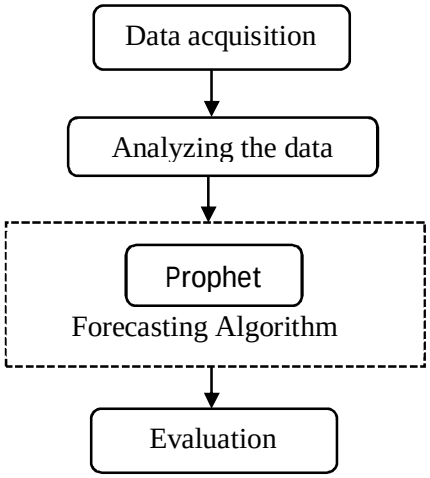


Figure 8 Time series forecasting architecture

C. Prophet Forecasting Algorithms

The Prophet is a straight-forward algorithm to develop the forecasting model for time-oriented data that offers sensible and more prominent forecasts. In 2017, this algorithm was developed by Facebook's Core Data Science Team which can be implemented in both python and R. This model depends on the additive regression approach that identifies trends, season, holidays and then consolidates them together using the eqn. 1.

Forecasting = g(t) + s(t) + h(t) + e(t) (1)

Prophet provides a) Saturation growth and b) piecewise linear model for g(t). Piecewise linear model is considered as default approach for forecasting and can be measured using the eqn (2). This model is flexible to both linear and non-linear functions and robust to outliers. Prophet is persistent in handling missing data [12] and can capture transitions in the trend. The accuracy of forecasting model will be as similar to the professional forecasters.

$$g(t) = (k + a(t)^T\delta)t + (m + a(t)\gamma) \quad (2)$$

Where k indicates growth rate, δ denotes rate adjustments, and m implies offset parameter. Also, seasonality model allows the frequency modifications such as daily, weekly, yearly, quarterly and this often helps in business time series analysis. Prophet uses Fourier Series to construct the model based on periodic as in eqn (3)

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)\right) \quad (3)$$

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VIII Aug 2020- Available at www.ijraset.com

Algorithm: Prophet forecast

1) *Step 1:* The prophet algorithm mainly depends on two input parameters (a) ***ds*** (Datestamp) column - ought to be in the desired format YYYY-MM-DD as demanded by pandas and (b) *y* column - incorporates only the numeric values taken into consideration for forecasting the future value.
2) *Step 2:* To extend the ds column with a specified number of dates *makes_future_dataframe* method should be applied.
3) *Step 3:* Specify the total number of days in *periods* parameter.
4) *Step 4:* Now, invoke the *predict* method that produces each row with the date and corresponding predicted values 'yhat'. Along with yhat, the objects contain the lower bound (yhat_lower) and upper bound (yhat_upper).
5) *Step 5:* The *Plot* function is applied to visualize the forecast.
6) *Step 6:* Now, the *cross_validation ()* is assigned to the model with *initial* and *horizon* parameters.
7) *Step 7:* Finally, *performance_**metrics*** is used to obtain various evaluation metrics for each horizon by comparing actual value and forecasted value (yhat).

V. EXPERIMENTAL ANALYSIS

In this paper, the most prominent forecasting algorithm is deployed to forecast daily cases in COVID. Table 1 illustrates the sample data considered for forecasting. For performance metrics, we have computed Root Mean Squared Error, Mean Squared Error, MAPE. According to WHO, a total of 1.44 Million people was affected by the consequence of disease in India. As mentioned earlier, India had a total population of approximately 1 billion, out of which 0.05% of the people were affected. Also, the recovery rate is around 60.2% that is proportionately higher than any other country

Date	Confirmed	Recovered	Deaths
6/22/2020	440464	248138	14016
6/23/2020	456120	258600	14484
6/24/2020	472988	271689	14908
6/25/2020	491193	285672	15909

Table 1. Sample Dataset

In this research, we are concern about the confirmed cases of COVID in India. So, we assign ds column with timestamp and y data frame with confirmed cases in India as required by the prophet. Table 2 demonstrates the transformation of the dataset as requisite by the algorithm.

ds	y
6/22/2020	440464
6/23/2020	456120
6/24/2020	472988
6/25/2020	491193

Table 2. Forecasting table

Once the dataset is fit into the model, then the next process is to predict the values. Table 3 illustrates the predicted values along with upper bound and lower bound.

ds	yhat	yhat_lower	yhat_upper
2020-11-25	3509104	1014729	5911417
2020-11-26	3528850	1017432	5944769
2020-11-27	3548597	1020474	5995134
2020-11-28	3568344	1007530	5995134
2020-11-29	3588090	9922960	6039108

Table 3. Predicted values

©IJRASET: All Rights are Reserved

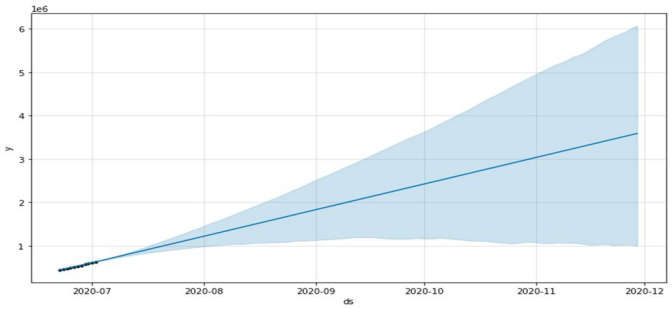


Figure 9. Forecasting

The accuracy of the prediction is essential to resist the critical circumstances and to save on time and cost. The *cross_validation* method is applied to measure the error using actual data. This method can be accomplished by selecting the cut-off value in the original data and fitting the model using this sample until the cut-off points.

	Horizon	mse	rmse	mape	Coverage
0	1 days	2637986.0	1624.0	0.002	0.0
1	2 days	10519946.0	3243.0	0.005	1.0
2	3 days	13613926.0	3690.0	0.006	1.0
3	4 days	2617808.0	1618.0	0.002	1.0

Table 4. Evaluation metrics

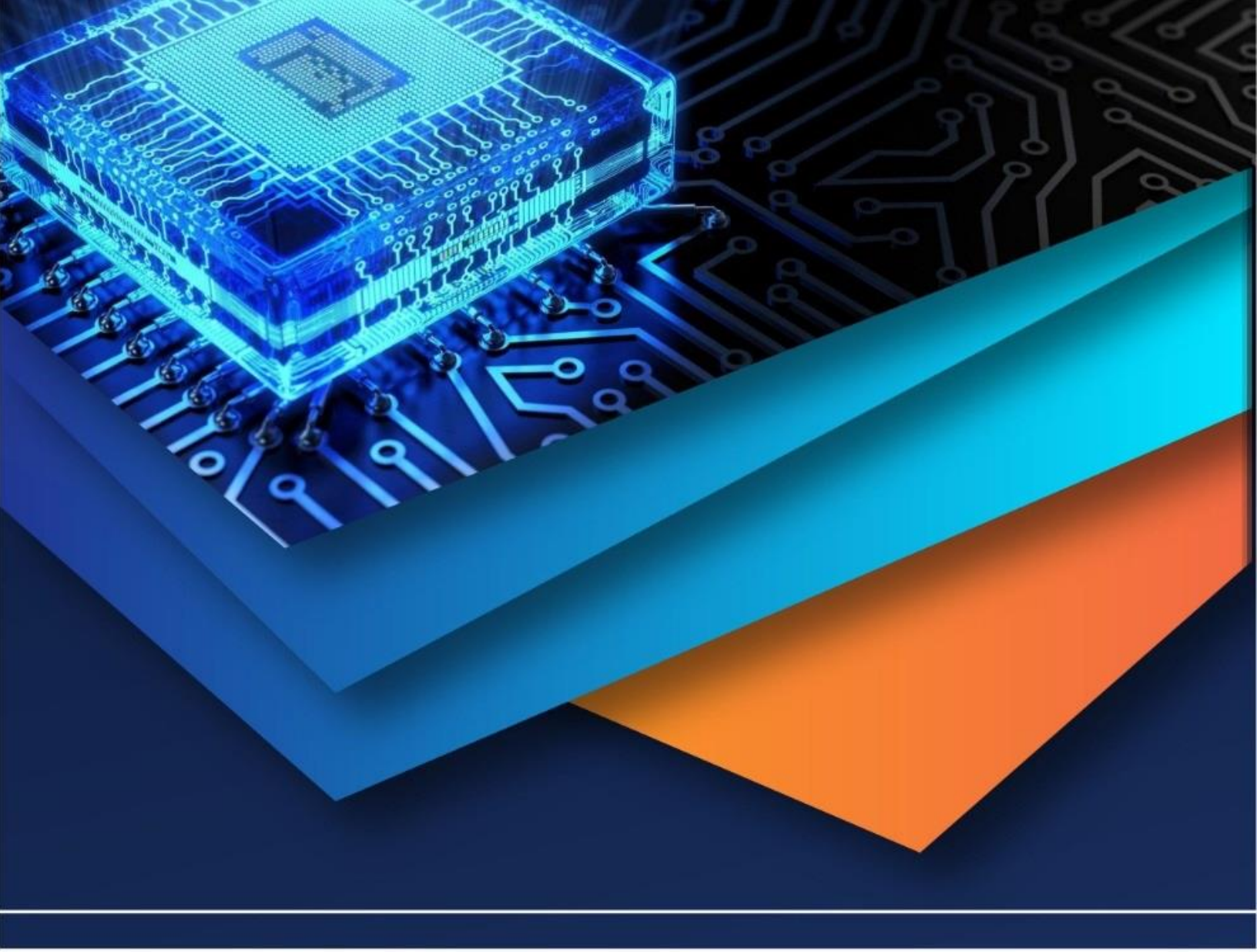
VI. CONCLUSION

In our research, we audited forecasting techniques and conducted experiments on the COVID-19 dataset from a reliable repository. This investigation is to forecast the COVID cases in India using the Fbprophet model using the available data associated with daily confirmed cases in India. Based on this, the daily forecast is around 35,88,090 cases per day for the next 120 days. This forecast is done with prevailing conditions. However, it can be enhanced by taking a few precautionary steps. This model can be further improvised with the upcoming data. For the future study, the model can be modified to predict the scenario of confirmed, deceased, and recovered instances of COVID-19.

REFERENCES

[1] Yunus, K., Thiringer, T., & Chen, P. (2016). ARIMA-Based Frequency-Decomposed Modeling of Wind Speed Time Series. IEEE Transactions on Power Systems, 31(4), 2546–2556.
[2] Yanchun Pan, Mingxia Zhang, Zhimin Chen, Ming Zhou, & Zuoyao Zhang. (2016). An ARIMA based model for forecasting the patient number of epidemic disease. 2016 13th International Conference on Service Systems and Service Management.
[3] Yenidogan, I., Cayir, A., Kozan, O., Dag, T., & Arslan, C. (2018). Bitcoin Forecasting Using ARIMA and PROPHET. 2018 3rd International Conference on Computer Science and Engineering (UBMK).
[4] Fang, W.-X., Lan, P.-C., Lin, W.-R., Chang, H.-C., Chang, H.-Y., & Wang, Y.-H. (2019). Combine Facebook Prophet and LSTM with BPNN Forecasting financial markets: the Morgan Taiwan Index. 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS).
[5] Amei Amei, Wandong Fu, Chih-Hsiang Ho, "Time Series Analysis for Predicting the Occurrences of Large-Scale Earthquakes", Journal of Applied Science and Technology, Vol 2, pg: 64-75, 2012.
[6] Ray, M., et al., Technology forecasting using time series intervention based trend impact analysis for wheat yield scenario in India, Technol. Forecast. Soc. Change ,2017.
[7] Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, et al. "The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China", International Journal of Infectious Diseases. Vol.91, Pg. 264–266, 2020.
[8] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study", Lancet, Vol:395, No:10223, Pg: 507–513, 2020.
[9] e Q, Wang B, Mao J, "The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19". J. Infect, Vol.80, No. 6, Pg: 607–613, 2020.
[10] Li G, De Clercq E , "Therapeutic options for the 2019 novel coronavirus (2019-nCoV)", Nature Reviews. Drug Discovery, Vol. 19, No. 3, pg: 149–150, 2019.
[11] Stadnytskyi V, Bax CE, Bax A, Anfinrud P, "The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission". Proceedings of the National Academy of Sciences of the United States of America, vol.117, No.22, Pg:11875–11877, 2020.
[12] Taylor SJ, Letham B. 2017. Forecasting at scale. PeerJ Preprints 5: e3190v2.
[13] Emir Zunic et-al, "Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data", International Journal of Computer Science & Information Technology, vol 12, Pg: 23-36, 2020.
[14] T. Chadefaux, "Early warning signals for war in the news", Journal of Peace Research, Vol. 51, No.1, pg: 5-18, 2014.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)