



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IX **Month of publication:** September 2022

DOI: <https://doi.org/10.22214/ijraset.2022.46897>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Covid-19 End and Peak Prediction Using Machine Learning

Sai Praveen¹, Ravi Kant Kumar²

¹Department of Computer Science and Engineering, SRM University – AP, Andhra Pradesh, India

Abstract: *The transmission of COVID-19 disease has shown a great impact on society. The whole world has been fighting this epidemic since late February 2020. The main objective of this project is to predict the spread and end of the COVID-19 disease. Because of the COVID-19 outbreak, the world's economy has been affected so far and an accurate prognosis of its epidemic is significant. Prediction for the end of this disease is not an easy task as it requires plenty of data and also various parameters involved in the prediction. This project proposes certain machine learning techniques and ARIMA models with numerical approximations from the dataset provided we forecast the number of reported cases and disease transmission.*

Keywords: *ARIMA Model, Visualization, Seasonal Decomposition, SARIMA Model*

I. INTRODUCTION

COVID-19 has affected every country around the world. The number of infected and dead people due to this disease has been raised. Prediction models contribute knowledge of the disease and its prevalence. These strategies examine past occurrences and scenarios to produce the best predictions for the future. These forecasts may aid people in preparing for potential outcomes[1][2]. They play a major role in getting accurate predictions. These models can either be predicted by mathematical models or by machine learning techniques. Data should be collected from different sources in large quantities to prepare an accurate model. Various parameters such as environmental factors, quarantine period, disease spreading rate, immunity levels of the person, person's past health issues, etc., are taken before forecasting the pandemic[5]. People throughout the world are researching many different ways to end this disease and are also predicting models using machine-learning techniques[6].

The model is built to forecast the number of confirmed cases, recovered cases, and death cases based on the data available. For the prediction model, the time series forecasting method is applied.

II. METHODOLOGY

Many mathematical models are used to predict the graph trend, but we used the ARIMA model for predicting the virus graph for the coming days. The ARIMA model is applied to forecast the data, which is high in terms of accuracy. ARIMA (Auto-Regressive Integrated Moving Average) model is a forecasting algorithm, the information contained in the past values of the time series can be utilized to forecast future values on its own. It is a type of model that describes a time series based on its previous values. As a result, that equation can be used to predict future values. This model is used to gain a better understanding of the data and to forecast future points in the series. It is applied to time series forecasting and provides complementary methods to the problem.

Prediction models contribute knowledge of the disease and its prevalence. These strategies examine past occurrences and scenarios to produce the best predictions for the future. These forecasts may aid people in preparing for potential outcomes. They play a major role in getting accurate predictions. These models can either be predicted by mathematical models or by machine learning techniques. Data should be collected from different sources in large quantities to prepare an accurate model. Various parameters such as environmental factors, quarantine period, disease spreading rate, immunity levels of the person, person's past health issues, etc., are taken before forecasting the pandemic.

III. RELATED WORK

1) *AR Model:* It is a statistical model which predicts future value based on its past values is called lags. The model that depends only on one lag in the past is given below:

$$Y_t = \omega + \phi Y_{t-1} + e_t$$

This model is called the long-memory model. If the recursion in time goes back until the beginning of the series, those are called long memory models.

- 2) *MA MODEL (Moving Average Model)*: It is a model which predicts the future based on past errors called errors. It depends only on the lag of error in the past. The model that depends only on one lag in the past:

$$Y_t = \omega + \theta e_{t-1} + e_t$$

These are called short memory models. There is no effect on the present predicted value if there are any big errors long enough ago.

Arima Model: Auto-Regressive Integrated Moving Average Model ARIMA model is a forecasting algorithm, the information contained in the past values of the time series can be utilized to forecast future values on its own. It is a type of model that describes a time series based on its previous values. As a result, that equation can be used to predict future values.[3] This model is used to gain a better understanding of the data and to forecast future points in the series. It is applied to time series forecasting and provides complementary methods to the problem. An ARIMA model is characterized by 3 terms:

- p, d, and q.
- 1. p - order of AR.
- 2. q - order of MA.
- 3. d - differencing we did for our time series stationary.
- 4. AR term - It is based on past values.
- 5. MA term - It is based on past errors.

$$Y_t = \omega + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t$$

The first step to build an ARIMA model is to make the time series stationary because 'Auto Regressive' in ARIMA means it is a linear regression model that uses its lags as predictors.[4] Linear regression models work best when the predictors are not correlated and are independent of each other.

What is p?

p is the order of the AR (Auto-Regressive) term. It refers to the number of Y lags to be utilized as predictors.

What is q?

q is the order of the MA (Moving Average) term. It refers to the number of lagged forecast errors that should go into the ARIMA Model.

What is d?

The most common approach to make the series stationary is to differentiate it i.e, subtract the previous value from the current value depending on the complexity of the series, more than one differencing may be needed. The value of d is the minimum number of differences needed to make the series stationary. If the time series is stationary, then d = 0.

- 3) *SARIMA Model*: If a time series exhibits seasonal patterns, seasonal terms must be included, becoming SARIMA, short for 'Seasonal ARIMA'.

- 4) *Seasonal Order*
(P,Q,D,Seasonality)

IV. VISUALIZATION

The CSV file contains the number of confirmed, active, and recovered cases along with the number of deaths. The file was updated with figures on 5th, July 2021. The file can be downloaded from[9]. You can also get the total data on COVID-19 in India[8]. And data for the total world can be found on WorldOmeters[10].

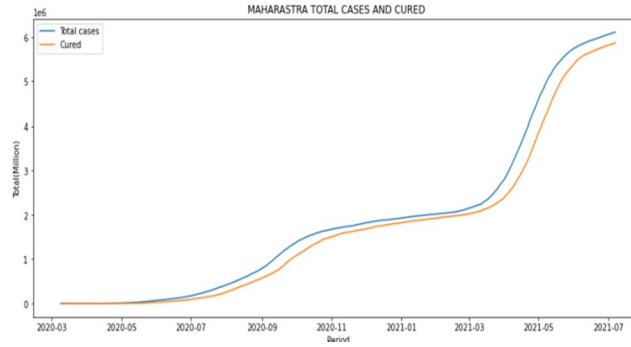


Figure 1: Visualization of the total cases and cured in Maharashtra

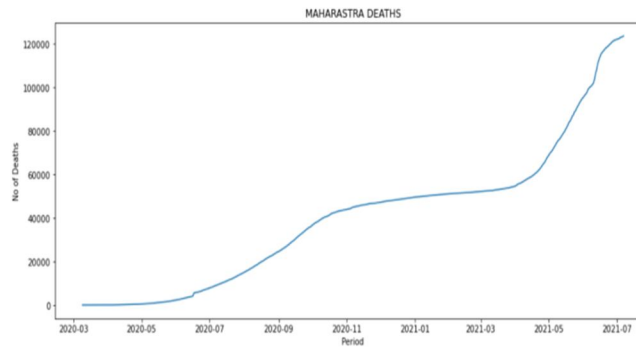


Figure 2: Deaths raised in July 2020 and in April 2021

Figure 3 indicates the number of total cases between March 2020 to July 2021. This gives a view of how the total cases were raised.

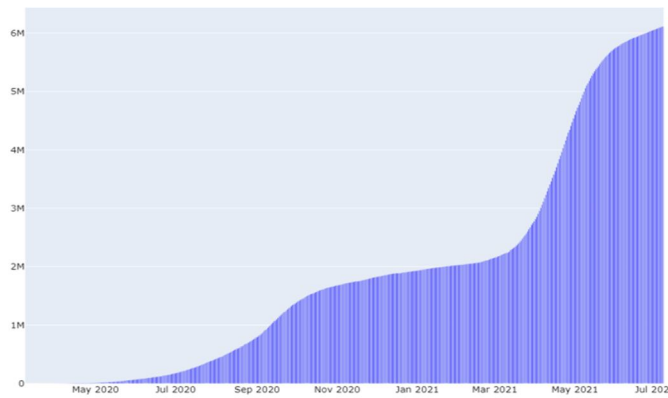


Figure 3: Graph indicating the raise of cases

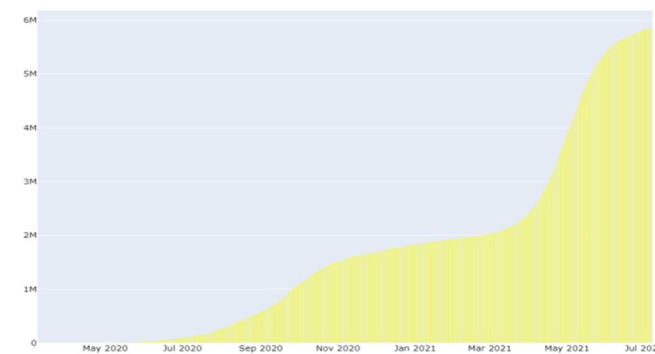


Figure 4: Graph interpreting the number of people cured at a high rate.

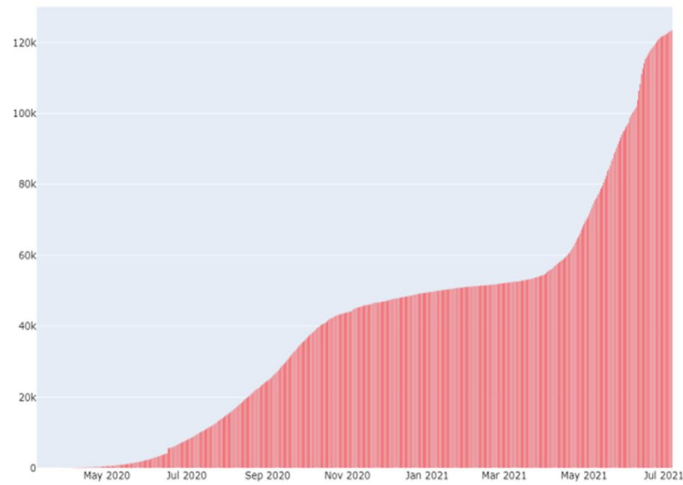


Figure 5: Graph indicating the number of people deceased.

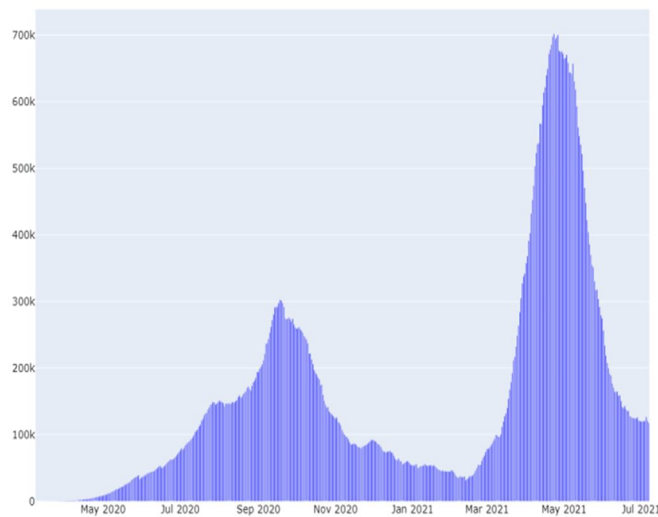


Figure 6: Graph indicating the total active cases.

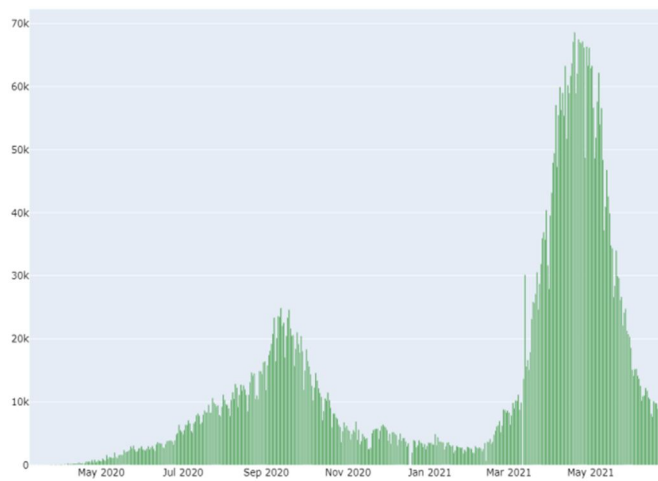


Figure 7: Graph indicating the 'daily new cases'

A. Lockdown Analysis

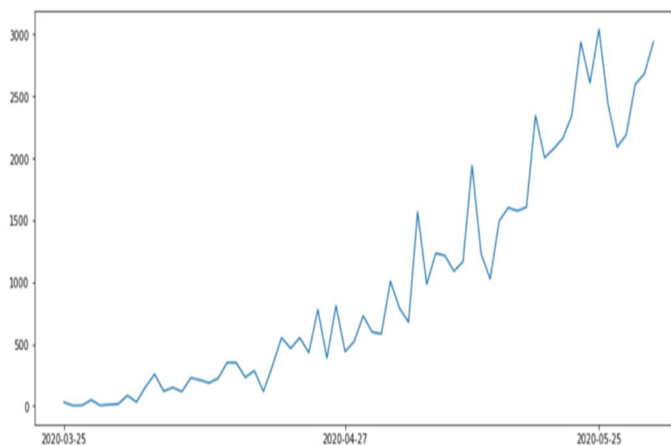


Figure 8: Lockdown 1

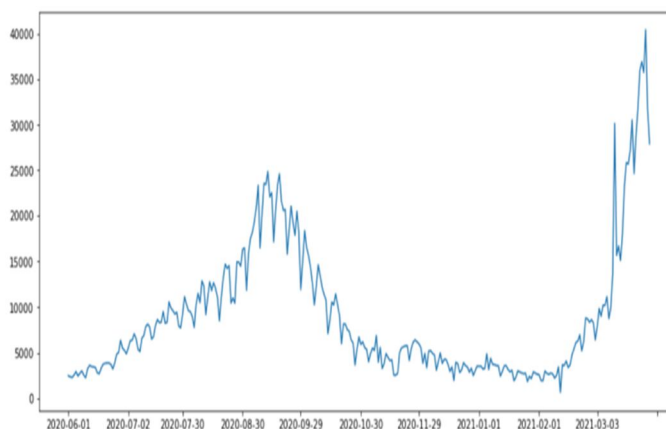


Figure 9: Unlock

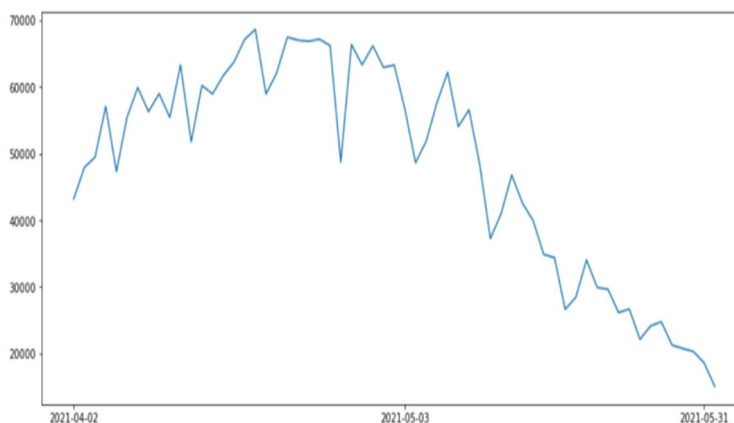


Figure 10: Lockdown 2

V. IMPLEMENTATION

We used the ARIMA Model for the prediction of the “number of daily cases “. We took 4 lags and 5 errors by choosing the best values for p and q using a for a loop. Checked Stationarity by Dickey-Fuller Test (ADF test). Our predicted values depend on its past 4 lags and 5 errors with 7 days of seasonal component.

A. Checking Stationarity

To check stationarity we used the Dickey-Fuller test. Dicker Fuller Test is a common statistical test used to test whether a given time series is stationary or not.

Here, the p-value is less than 0.05 so our data is stationary. P-value > 0.05 , our data is not stationary. P-value < 0.05 , our data is stationary[3][4].

B. Seasonal Decomposition

Seasonal decomposition involves thinking of a series as a combination of level, trend, and seasonality.[4]

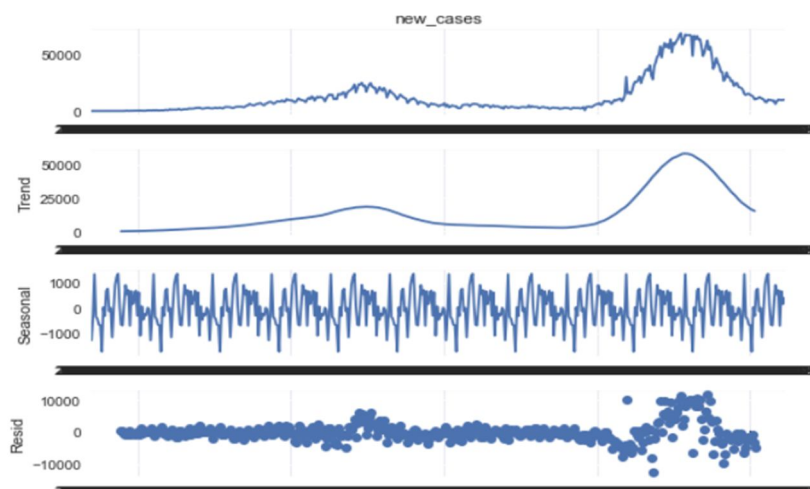


Figure 11: Analysis to summarise (or, occasionally, combine) three elements: the linear trend, the periodic (seasonal) component, and the random residuals.

C. ACF

The coefficient correlation between two values in a time series is called the AutoCorrelation Function. We select the MA using AutoCorrelationFunction. Here are the best MA values 1 and 5[4].

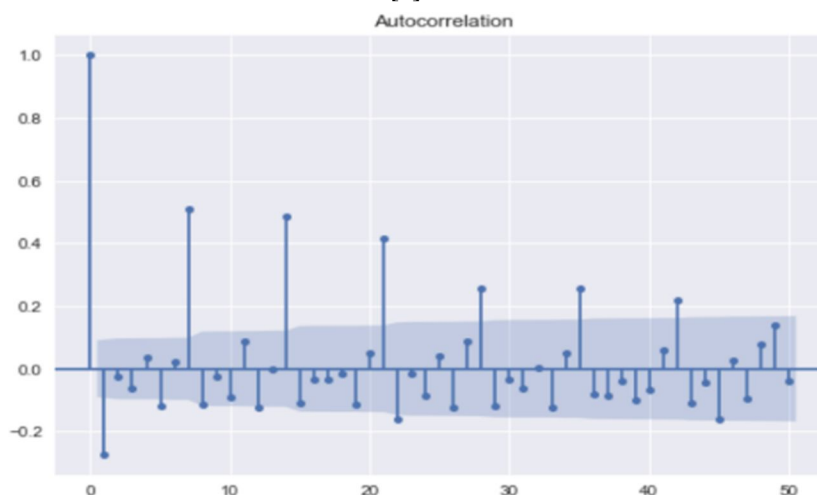


Figure 12: the average relationship between the data points in a time series and the data points before them.

D. PACF

Partial autocorrelation is the summary of the relation between a value with value at previous time steps. We select AR with Partial autocorrelation.

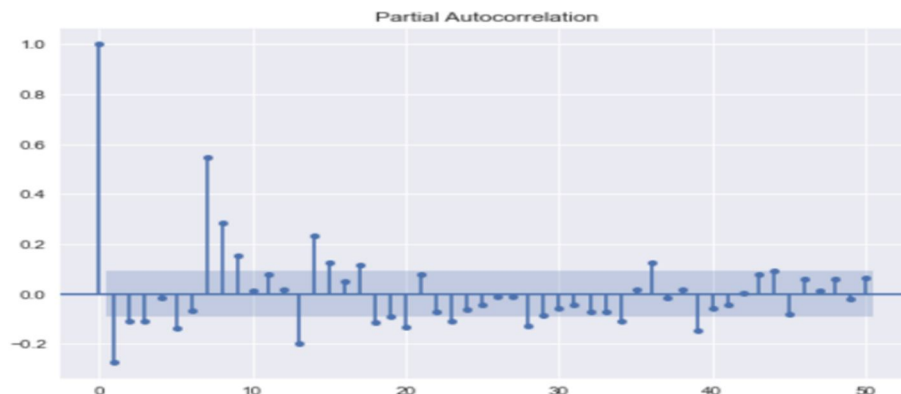
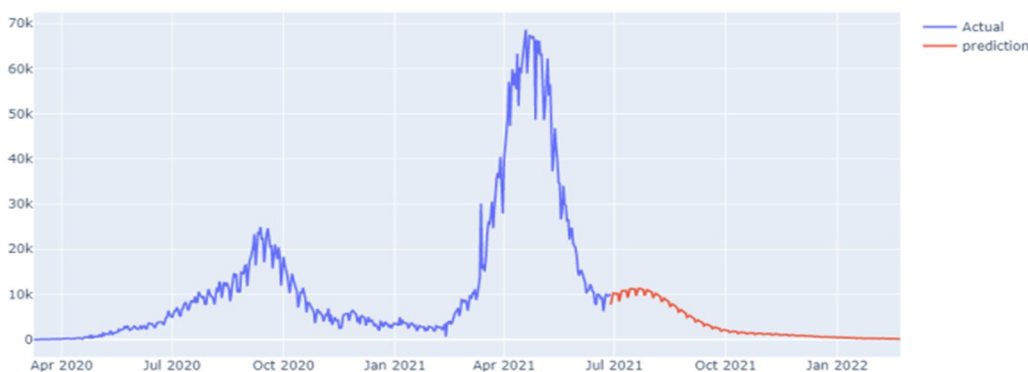


Figure 13: Regresses the values of the time series at all lesser lags and shows the partial correlation of a stationary time series with its own delayed values.

To find the d , we used `ndiffs` from Arima. Firstly, we should fit the training set with the SARIMAX model. Then we need to predict the test data. So, we used `summary_frame` for getting the data frame of predicted values.

VI. RESULT

In the above study, we have assessed the situation of Covid-19 in India. We have developed equations from March to June for the number of cases, deaths, and active cases. We then used the regression model to forecast the cases for upcoming days along with the accuracy percentage. The results are also provided in the above section. The project can be further developed by taking many other factors like vaccinated people, vaccination rate, variants and their spread rate, the measures taken like lockdown, etc. This mathematical model of forecasting cases for the future will surely prove to be helpful in the fight against the pandemic.



REFERENCES

- [1] Tiwari, Sunita, Sushil Kumar, and Kalpna Guleria. "Outbreak trends of coronavirus disease–2019 in India: a prediction." *Disaster medicine and public health preparedness* 14, no. 5 (2020): e33-e38.
- [2] Jia, Lin, Kewen Li, Yu Jiang, and Xin Guo. "Prediction and analysis of coronavirus disease 2019." *arXiv preprint arXiv:2003.05447* (2020).
- [3] Benvenuto, Domenico, et al. "Application of the ARIMA model on the COVID-2019 epidemic dataset." *Data in brief* 29 (2020): 105340.
- [4] Hillmer, Steven Craig, and George C. Tiao. "An ARIMA-model-based approach to seasonal adjustment." *Journal of the American Statistical Association* 77.377 (1982): 63-70.
- [5] Zhong, Linhao, Lin Mu, Jing Li, Jiaying Wang, Zhe Yin, and Darong Liu. "Early prediction of the 2019 novel coronavirus outbreak in the mainland China based on simple mathematical model." *Ieee Access* 8 (2020): 51761-51769.
- [6] Bentout, Soufiane, Abdennasser Chekroun, and Toshikazu Kuniya. "Parameter estimation and prediction for coronavirus disease outbreak 2019 (COVID-19) in Algeria." *AIMS Public Health* 7, no. 2 (2020): 306.
- [7] Abdi, Milad. "Coronavirus disease 2019 (COVID-19) outbreak in Iran: Actions and problems." *Infection Control & Hospital Epidemiology* 41, no. 6 (2020): 754-755.
- [8] 2021, Worldometer, Coronavirus Data in India. Available <https://www.worldometers.info/coronavirus/country/india/>
- [9] Covid-19 India Data Set, Ministry of Health and Family Welfare, Govt of India. Available <https://prsindia.org/covid-19/cases>
- [10] Covid-19 India Total Data, Covid19IndiaOrg. Available <https://www.covid19india.org/>



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)