

Covid-19 Time Series Forecasting

Submitted to:- Dr Sanjay Saxena



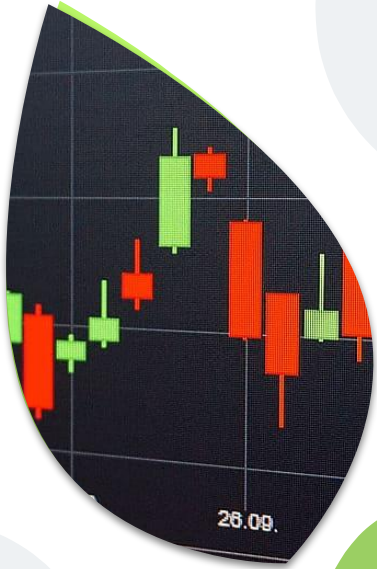
Objective of our Project:

This study aims to predict the number of confirmed individuals, vaccination and deaths by the coronavirus. The research is mainly focused on predicting time-series data using Arima and LSTM model. The result helps the health institutions and the public on the understanding of the impacts and influence of the coronavirus in their life and the importance to control its spread.

Time series data is a collection of observations obtained through repeated measurements over time. Plotting the points on a graph one of the axes would always be time.



Background of our work:



The novel Coronavirus, COVID-19 is a threat to the whole world. It has impacted economies, and the livelihood of countries and the people in them.

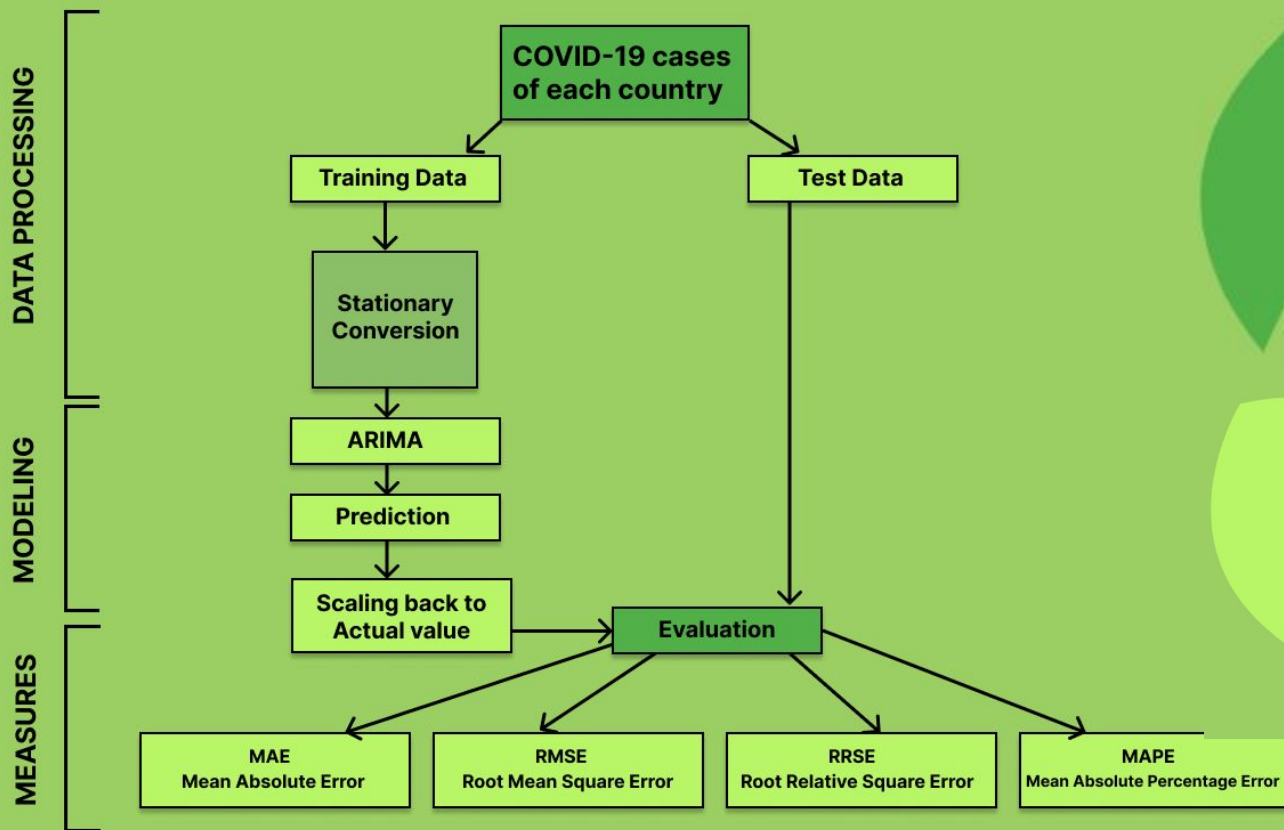
Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data

Time series forecasting is the use of a model to predict future values based on previously observed values.

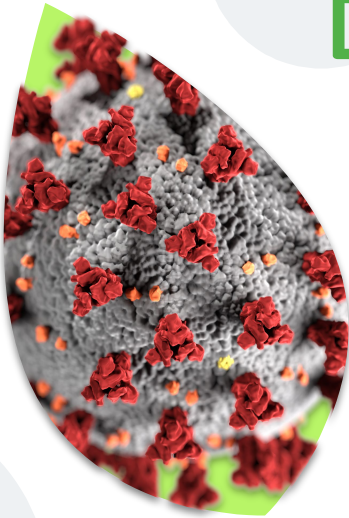
There are different types of time series forecasting models:-

- ARIMA
- Moving Average model
- Exponential smoothing Model

Flow of Development:



Data Acquisition and Preparation



“

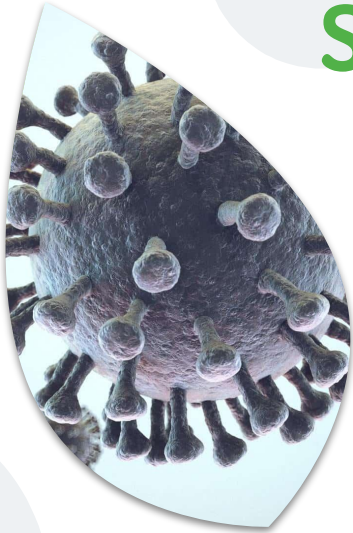
Data Source:

Here we are using data provided through API
'http://api.covid19india.org/states_daily_csv/confirmed.csv' .

We are importing this CSV file using pandas and setting index column as 'date' because the values will be forecasted according to this values.



Steps for Data Analysis:



- “★ the data analysis phase includes first importing necessary libraries for analyzing the dataset like NumPy, Pandas, matplotlib, Timedelta and DateTime.
- ★ these libraries help the programmer to analyze any data. As the dataset is in .csv format, we import the dataset.
- ★ we first read the dataset and then store it in pandas dataframe.



ARIMA for Time series Forecasting:

“ The Autoregressive Integrated Moving Average (ARIMA) method, It combines both Autoregression (AR) and Moving Average (MA) models as well as a differencing pre-processing step of the sequence to make the sequence stationary, called integration (I).

- ★ AR: Autoregression. A model that uses the dependent relationship between an observation and some number of lagged observations.
- ★ I: Integrated. The use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.
- ★ MA: Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

ARIMA for Time series Forecasting:



ARIMA takes two arguments

First is the column whose value is to be forecasted. In our case it is 'TT' column of our data frame.

Second argument is a tuple named 'order'. It takes three values :

- ★ *p: The number of lag observations included in the model, also called the lag order.*
- ★ *d: The number of times that the raw observations are differenced, also called the degree of differencing.*
- ★ *q: The size of the moving average window, also called the order of moving average.*

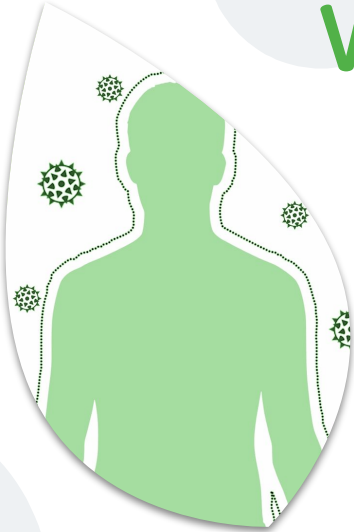


Performance Metrics:



- “ ★ MAE is used to determine how close the forecasted values and eventual outcomes are.
 $MAE = \frac{\sum_{i=1}^N (z_i - \hat{z}_i)}{N}$
- ★ RMSE is defined as standard deviation of the differences between values predicted and the values observed. $RMSE = \sqrt{\frac{\sum_{i=1}^N (z_i - \hat{z}_i)^2}{N}}$
- ★ MAPE is another measure of prediction accuracy of a forecasting model. It is defined as: $MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|z_i - \hat{z}_i|}{z_i}$
- ★ RRSE is relative to what it would have been if a simple predictor had been used.

Working Environment:



“



Jupyter Notebook: An open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text.



Ram- 16GB



GPU- T4 x2



Language- python 3.11.2



Run time GPU- 13 minutes



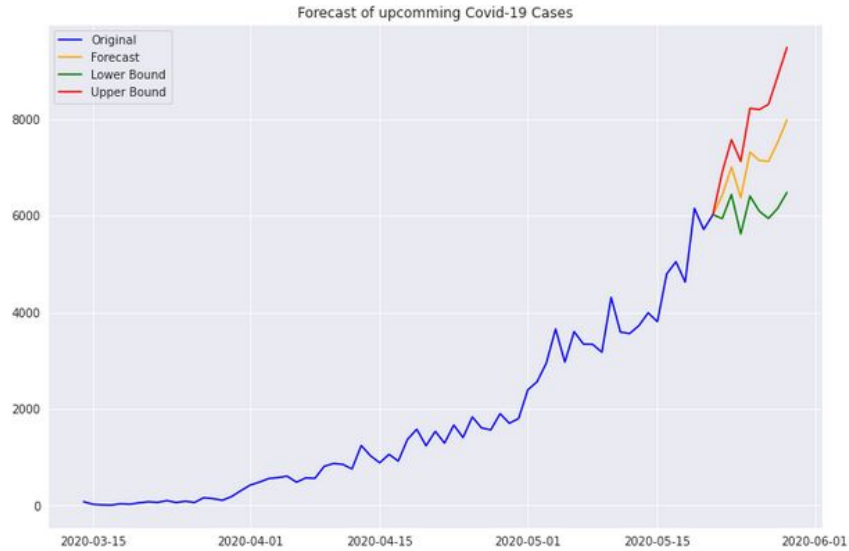
Version controller System- GIT



Repository hosting- GitHub



Results Obtained:





Discussion:

“

- ★ The result of prediction is stored in variable 'yhat'.
- ★ Displaying content of 'yhat' shows date and prediction. Hence according to our model, around 938.55 cases will be reported on 2020-04-16.
- ★ The Akaike Information Criteria (AIC) is a widely used measure of a statistical model. ... When comparing two models, the one with the lower AIC is generally “better”.
- ★ We will use plot_predict method which will also plot graph for us along with predicting values.
- ★ We are here using datetime library of python. Start date is set to today . Timedelta is used to do addition and subtractions on date and times. Here we have declared an object td as timedelta with argument date=7. start_date+td gives us 2020-04-16+ 7 days = 2020-04-23

Conclusion:

There are many alternative models from which we can perform the time series forecasting, but ARIMA is pretty simple to comprehend. Time series forecasting is really useful when we have to take future decisions or we have to make analysis.



Future Scope:



- As the pandemic progresses, more accurate and comprehensive data becomes available, which can enhance the accuracy of our model.
- With the emergence of new variants of the virus, it becomes crucial to monitor and forecast their spread. Our model can be adapted to detect outbreaks, track variant-specific trends, and provide early warnings for potential hotspots.
- It can assist policymakers and public health officials in evaluating the impact of interventions, such as lockdown measures or vaccination campaigns.
- Accurate forecasts can help hospitals, healthcare systems, and governments plan for future demand for medical resources, such as hospital beds, ventilators, and medical supplies.
- Our model can be combined with other ML techniques to further improve the forecasting.

Role of the members:



- **Data Processing**– Smita Tarai (B420050)
- **Modeling**– Sakshi Singh (B420044)
- **Measures**– Samrudhi Panda (B420045)



Thank You!